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

Joshua S. Hendrick
Thomas J. Howell
Darrin M. London
Eric M. Luehrs
Michael Saliba
Donald Brown
Jason Dalton
Francis B. Prats

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

Department of Systems and Information Engineering
University of Virginia
brown@virginia.edu

ABSTRACT

Last year, a team in the Department of Systems Engineering at the University of Virginia developed a web-based crime analysis toolkit (WebCAT). This web-based system allows for information sharing and cross-jurisdictional analysis between the various law enforcement agencies throughout Virginia. However, this system was not deployed. This year’s capstone team has improved WebCAT’s functionality, robustness, and quality of analysis to enable deployment. This work has been done in the context of the whole systems lifecycle: (1) analysis; (2) design; (3) integration; (4) testing; and (5) evaluation. The capstone team has added a number of components to improve data sharing and crime analysis. These new components to WebCAT, among others, have improved the capabilities of crime analysts in terms of (1) quality of analysis; (2) increased speed of performing an analysis; (3) minimal cost necessary to design, implement, use, and maintain the application.

1 INTRODUCTION

Crime prevention and response is an essential element in providing a safe environment for all citizens in the state of Virginia. Information sharing and crime analysis have proven very effective in achieving this goal. (VCAN Conference, 2004) Unfortunately, very few local law enforcement agencies in the State of Virginia have the capability for cross-jurisdictional sharing of data, and most lack even basic resources to perform in-depth crime analyses.

WebCAT is an open source application that allows law enforcement agencies throughout Virginia to share information and perform cross-jurisdictional crime analysis. The following document reports on the results from this year’s capstone team and shows how the team has improved WebCAT’s functionality, robustness, and quality of analysis. Specifically, the team has added or improved the following elements:

1. Geographic Information System (GIS) Mapping Application
2. Incident Matching Capabilities
3. Administration Tools Functionality
4. User Interface

Incident matching capabilities have been implemented to identify similar crime incidents using an advanced search algorithm that optimizes precision and recall. Using the new GIS mapping system, analysts can query crime incident data, plot incidents on the fly, and retrieve information for a specific incident. Additionally, the WebCAT application is scheduled for formal testing and evaluation at the end of April in preparation for its pilot test in the Richlands, Virginia police department.
2 CURRENT STATUS OF CRIME ANALYSIS

2.1 Crime Analysis in Local Jurisdictions

Every police department in the state currently does crime analysis differently. Most of these local law enforcement agencies do not have the technical expertise to formulate an in-depth crime analysis. Many others simply do not have the necessary human resources to devote to this task.

Each local jurisdiction performs their own crime analysis, which produces a wide range of results depending on the jurisdiction’s resources and technical expertise. Most analysis is limited to basic summary reports that identify monthly or yearly trends. A majority of the police departments do not have access to computer-generated map software that plots the locations of crimes in their jurisdiction. Limited querying capabilities are available, but the results of the query are not meaningful without the proper tools of turning this data into useful information for decision-making.

The limitations in local crime analysis are mainly due to a lack of (1) personnel (2) time (3) accurate and complete data, and (4) funding.

2.2 Statewide Crime Analysis

The Virginia Department of Criminal Justice Services (DCJS) performs statewide crime analysis. DCJS has just one employee on its crime analysis staff, who identifies monthly and yearly trends in crime data around the state. DCJS reports the following problems: (1) lack of standard practices in reporting (2) lack of accurate and complete data, and (3) “limited to cumbersome” crime analysis capabilities. The problems reported by DCJS and local law enforcement agencies re-affirm the need for a powerful and consistent crime analysis tool at a local and statewide level.

2.3 Crime Reporting Format

All effective criminal analysis techniques require accurate, complete, and updated crime data. Within Virginia, data collected at each local jurisdiction are sent to the Virginia State Police, who in turn send the aggregated data to the FBI in the National Incident Based Reporting System (NIBRS) format, which contains roughly fifty crime characteristic fields to describe an incident. However, data collected by the state are not readily available to the local agencies. As a result the local agencies are unable to easily perform cross-jurisdictional crime analysis.

3 OVERVIEW OF WEBCAT DESIGN

WebCAT addresses the limitations reported in section 2 by allowing law-enforcement agencies to share information and perform cross-jurisdictional crime analysis. Furthermore, WebCAT is web-based and comprised of mostly open source technology, making it a low cost solution. Extensible Markup Language (XML) technology is used to allow users to perform an analysis on local, regional, and statewide data. The capstone group created a goals tree at the start of the year to outline the high level WebCAT design goals and all subsequent tasks required to accomplish those goals. After providing the rationale for the specific improvements to the WebCAT design, the capstone team followed the systems approach for each new component added and each existing component which was improved upon.

First, the group analyzed viable solutions to the stated high level goal and selected the most feasible alternative. Next, the team undertook the design phase of the project which focused primarily on coding using web programming technologies which included, but were not limited to: XML, Active Server Pages (ASP), and JavaScript. The team also drafted functional and conceptual requirements for each task. Next, each individual component was integrated into the existing WebCAT system. Finally, the team performed usability testing to assess the functionality and design of WebCAT. The success of each individual tool, and the aggregate of the components, was judged by the (1) increased speed of performing an analysis; (2) improved quality of analysis; (3) minimal cost necessary to design, implement, use, and maintain the application. The following subsections detail the specific work done for each component. This work followed the systems approach and was necessary to accomplish the stated goals for that task.

3.1 GIS Mapping Capabilities

3.1.1 Goal

The goal of the GIS application is to improve the visual representation of crime incident data by integrating a more powerful and robust GIS mapping solution.

3.1.2 Rationale and Background

Currently, only 47% of crime analysts in Virginia actually use GIS specific software in their analysis, according to results obtained from the 2004 Virginia Crime Analysis Network (VCAN) conference. Most crime analysts reported that they use thumbbacks on a map as well as graphs to visually represent data. Further, those who did use GIS software reported that it was not very user friendly and that
too much time is wasted having to manually input data. This year’s GIS application addresses these problems.

Last year, the WebCAT team implemented a Geographic Information System (GIS) using MapServer software. MapServer is an open-source GIS application that was developed by the University of Minnesota. This application only included a map of Tazwell County, Virginia. Therefore, users could only perform a visual analysis of data sets associated with Tazwell County. Users could map the data set on a Tazwell County map along with various layers for roads, streets, etc. However, if users wanted to visually analyze a new data set they would have to leave the GIS section in order to query the NIBRS database (Hawkins et al., 2003). This created an inconvenience for the user and an increase in the time needed to perform an analysis. The results of the Virginia Crime Analysis Network survey and the previous year’s limited GIS analysis capability demonstrated the need for a new application that enables analysts to perform statewide analysis using GIS. The requirements, as stated later in this section, improve an analyst’s ability to visualize crime incidents in their jurisdiction and around the state.

3.1.3 Systems Approach

This year’s WebCAT team decided that the MapServer software did not provide sufficient functionality for users. The group considered Manifold and Geo Media WebMap as possible GIS solutions for WebCAT. After analyzing both systems, the group found that Web Map had similar functionality to Manifold, but at a much higher licensing cost. Therefore, the team decided to use Manifold because of its web based capabilities, customizability, query functionality, and cost.

Next, the group defined the following requirements for the new GIS application: (1) allow users to query the NIBRS database directly from the GIS (2) provide users with map layers for the entire state of Virginia (3) enable users to retrieve specific information on a crime incident by clicking on the plotted point (4) integrate geocoding functionality to convert addresses to latitude and longitude values (5) integrate thematic mapping capability. In order to meet these requirements, the group needed to customize the Manifold software using JavaScript.

This year’s GIS application allows analysts to query the Virginia NIBRS database and view the results on a digital map. The application currently contains 6 individual county maps from the state of Virginia. However, this number can be expanded to include more counties, allowing WebCAT to be easily implemented in all jurisdictions in the state of Virginia. Once a user queries the database, an XML file is created in the user’s directory. An example XML file is shown below in Figure 1.

```xml
<?xml version="1.0" standalone="no" ?>
<crimes>
  <crime>
    <crim_no>VA0300002001001588</crim_no>
    <off_code>120</off_code>
    <date>Sat Sep 8 00:00:00 EDT 2001</date>
    <latitude>38.72</latitude>
    <longitude>-77.77</longitude>
  </crime>
  <crime>
    <crim_no>VA0300002001003890</crim_no>
    <off_code>120</off_code>
    <date>Sun Sep 8 00:00:00 EDT 2001</date>
    <latitude>38.72</latitude>
    <longitude>-77.77</longitude>
  </crime>
  <!-- More data... -->
</crimes>
```

Figure 1. XML Data Set

Once this XML file is created, the GIS application reads the file and maps the crime incidents on the corresponding county map. Figure 2 shows the result of a query of all robberies in Loudoun County.

![Figure 2. Result of Query in Loudoun County](Image)

The black dot in the center of the county is the resulting crime incident. A key limitation of this query is that only one point can be plotted because the NIBRS database does not store an address field for each crime. However, many
jurisdictions do track the addresses of incidents, but this field is then cut out by NIBRS. This is a problem that the state is going to address within the coming year.

The GIS application also allows analysts to use their own database to perform an analysis. If the database has an address field, the application can convert, or geocode, the address into a latitude and longitude. The team connected the application to a Richmond database that included an address field. Once these addresses are converted to values of latitude and longitude, the crime incidents are plotted on the map along with layers for streets, cultural points, schools, rivers, and highways. Users can then get more information about any incident or layer by clicking on that point, as depicted in Figure 3.

Figure 3. Incident specific data

Manifold also includes thematic mapping functionality. Thematic mapping shows the spatial distribution of one or more specific data themes for standard geographic areas. The map can be either qualitative or quantitative (Statistical Reference Centre, 2004). Using this feature, analysts can select a crime type for the state of Virginia. The counties in Virginia are then colored in different shades of red and yellow, corresponding to the amount of crime in that county. Those counties shaded in red have much higher amounts of crime incidents, while those in yellow have the lowest amounts.

Once the design of the GIS application was completed, the group integrated the system into WebCAT. Integration was successful despite problems arising due to differences in the programming languages used in WebCAT and the GIS application.

The capstone group recommends improvements to the GIS mapping tool in the form of kernel density layers. These layers would identify areas of high likelihood of future crimes based on the given dataset. This method of spatial analysis to identify volatile areas of crime is a very useful crime predicting method. It helps the analyst graphically identify geographically unstable areas of high crime.

Last years WebCAT application provided only limited functionality for users to visually represent crime incidents. This updated GIS application has improved the way crime analysts perform crime mapping.

In terms of Speed of Analysis:

- Users can query the NIBRS database and view the resulting incidents in less than a minute. For example, a query for all robberies in Loudoun County will take about 30 seconds.
- Enables users to convert addresses in a database to latitude and longitude values in less than thirty seconds. For example, 900 addresses can be converted to latitude and longitude in less than thirty seconds.

In terms of Quality of Analysis:

- Allows crime analysts to visualize the distribution of crimes in a certain area and instantly return specific information about any incident
- Uses thematic mapping to give analysts another visual representation of the number of crime incidents in different counties

3.2 Incident-Matching Query

3.2.1 Goal

The goal of the Incident Matching Query is to give analysts the capability to identify similar crime incident characteristics using an advanced search algorithm that optimizes precision and recall.

3.2.2 Rationale and Background

Incident matching algorithms are used to identify patterns and hidden relationships based on similarities in crime characteristics. A thorough incident matching tool can often lead to the swift apprehension of a criminal suspect. Law enforcement agencies in the State of Virginia currently use many different methods to perform incident matching. The most commonly used techniques are (1) Database queries (2) Customized Software (3) Human Recognition. (VCAN Conference, 2003)

Though two jurisdictions may be geographically adjacent to each other, it is difficult for most analysts to identify similar crime characteristics between the two because they do not have quick access to state-wide data. The most commonly used characteristics when searching for similar incidents in a crime data set, as obtained from the VCAN
Table 1. Percentage of agencies using specific crime characteristics (VCAN)

<table>
<thead>
<tr>
<th>Similarity Matching Criteria</th>
<th>Time</th>
<th>Location</th>
<th>MO</th>
<th>Victim Type</th>
<th>Criminal Type</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>81%</td>
<td>81%</td>
<td>75%</td>
<td>50%</td>
<td>72%</td>
<td>31%</td>
</tr>
</tbody>
</table>

WebCAT’s incident matching capability the team added uses a robust incident matching query, as described in section 3.2.3, and allows analysts to run a query with data from (1) within their own jurisdiction (2) across their region, and (3) across the state.

3.2.3 Systems Approach

The team selected an advanced incident matching search algorithm that optimizes precision and recall. Precision is the number of records returned that have significant similarities in crime characteristics to the search specifications. Recall is the total number of files returned by the search. The group chose to use the following pseudo algorithm which strikes a balance between precision and recall:

1. Determine a representative set of characteristics – These incident characteristics form the underlying data set in which comparisons are made. The initial incident matching algorithm contains those characteristics mentioned in Table 1 (time, location, criminal type, etc).
2. Apply weights to characteristics based on how likely a certain attribute will be in predicting similarities. For example, a match in blood DNA found at two sites is a much better predictor for similarity than a match in suspect hair color.
3. Apply the weight for every pairwise comparison to target. Apply the weight using either a Boolean or percentage criteria:
   a. Boolean: apply weight only if the pair are the exactly same. For example, if blood samples were found at the scene of two crimes but the DNA did not match, do not apply the weight. It is very unlikely that there is a relationship between these crimes based on this data.
   b. Percentage: there could still be a relationship between two incidents where different values were returned for a certain characteristic. For example, a suspect’s hair color can often be reported as brown when it was in fact black, and vice versa. In cases like these, a percentage of similarity will be determined for that characteristic and a weight will be applied to that percentage:
4. Identify Outliers-Finding similarities based on unusual characteristics. For example, if almost all weapons used are knives or pistols, and a very rare silver-plated pistol was used in two separate incidents, it is likely that there is a relationship between these two crimes.

Next, the group implemented the algorithm. This component of WebCAT was coded using XML, ASP, and VBScript. The crime dataset used is stored in an XML document which can contain local, regional, or state-wide crime incidents. The current incident matching application in WebCAT only makes use of steps 1 and 2 of this algorithm. Steps 3 and 4 of the algorithm must be added in the future, and are essential to robustness of the incident matching query.

The incident matching component was developed and coded using the templates and infrastructure of WebCAT, so it was easily integrated into the application. The incident matching component was included in the usability testing described in section 4 of this paper. Over 80% of users performed the task of identifying the location of the most similar incident successfully, and completed this task in an average time of 50 seconds.

Last year’s version of WebCAT did not include a functioning incident matching component. This new component has improved many crime analysts’ incident matching capabilities.

In terms of Speed of Analysis:
- Currently, there are different methods used by local law enforcement to perform incident matching as explained above. Depending on which technique is used, the analysis can take anywhere from one hour to several days.
- Using WebCAT, analysts are able to perform cross-jurisdictional incident matching analysis in less than 1 minute.

In terms of Quality of Analysis
- Allows users to search for similarities between individual criminal incidents in their jurisdiction, across their region, or across the state, using a high precision matching algorithm.
3.3 Administration Tools

3.3.1 Goal

The goal of updating the functionality of the administration tools was to provide a means for managing user profiles and administrative privileges in WebCAT.

3.3.2 Initial Status of Administration Tools

The capstone team has implemented the ability to Insert, Update, and Delete user accounts. At the beginning of this project’s lifecycle none of these commands were functional. The WebCAT application did however display the user profiles in the administration page as depicted in Figure 4.

![Figure 4. User Profiles Administration Page](image)

This administration page listed the following user information: (1) User ID (2) Password (3) First Name (4) Last Name (4) Administration Status, and (5) Operation. The user profiles page was developed using ASP and Extensible Stylesheet language (XSL) and dynamically retrieves user information stored in an XML document.

3.3.3 Improvements and Current Status of Administration Tools

An administrator who wants to perform an update, insert, or delete command must select an operation from the administration page and input data to the user form shown below.

![Figure 5. Administration Tools User Input Form](image)

Using this form the administrator is now able to (1) Insert a new user profile (2) Update information of a current user, and (3) Delete a user profile.

When the administrator chooses to insert a new user, he must fill in the entire user input form. There is error checking present to ensure that the profile is added according to the technical specifications of WebCAT, namely, (1) no field can be left blank; (2) email addresses must contain the @ symbol; (3) the status of the user must be set at a certain level. After the administrator successfully submits the insert form, WebCAT automatically updates the XML file, appending to it a new user record. WebCAT also creates a folder for the user where charts, graphs, and other analysis are saved and stored. The user name and password are activated and can now be used to successfully login to WebCAT.

From the administration page, the administrator also has the option of updating a user’s profile. Once changes are submitted, WebCAT automatically modifies the XML file, reflecting the changes for the selected user. Similar to the Insert User function, error checking is also present with the update command. The user ID field is disabled and serves as the permanent unique identifier of each user.

Finally, the administrator has the ability to delete user profiles. Once removed, the user record is automatically deleted from the XML document and the user’s folder is deleted from the WebCAT system. The user name and password combination can no longer gain access to the WebCAT system.

The administration tools component was included in the usability testing described in section 4 of this paper. Most test users were able to successfully create a new user profile and log in to WebCAT. As mentioned, this year’s capstone team has improved on the administration tools of the previous version of WebCAT by allowing administra-
tors to (1) insert a new user profile (2) update information of a current user, and (3) delete a user profile.

3.4 Graphical User Interface (GUI) Development

3.4.1 Goal

The goal of GUI development was to decrease the learning curve of WebCAT by creating an intuitive and simplistic interface.

3.4.2 GUI Rationale

The user interface communicates between the system and the user. From the user’s perspective, the interface is the system (Sauter, 2000). For this reason, a poor interface can lead to the ultimate failure of an application (Sauter, 2000). An effective, well-organized interface minimizes the user’s learning curve, is intuitive, eliminates confusion and allows the user to complete desired tasks effectively and efficiently.

3.4.3 Systems Approach

Three major changes were made to the user-interface of WebCAT: (1) changing the methodology for naming data-set files (2) creating a way to display what file the user is currently analyzing (3) graphically redesigning the user interface. Every change made to the GUI of WebCAT was motivated by the goal of creating a system that any user could use accurately, regardless of their technical background.

When an analyst queries the Virginia NIBRS database, an XML file is created containing the resulting crime incidents. Previously in WebCAT, every time the VA IBR database was queried, the new dataset overwrote the old dataset, meaning the analyst could only work with one dataset at a time.

Furthermore, the file name automatically created was based on the user’s identification tag, and provided no context about the contents of the file.

Now, when a user searches the NIBRS database, the query results are saved in an XML file that is named based on the location for the search and the type of crime. For example, if a crime analyst searches the database for robberies in Appomatox Co., the result of this search is automatically saved in a XML file labeled “Appomatox Co_Robberies.XML.” All query results can be viewed and selected by going to the File Directory function within WebCAT. Now, the user can maintain several different datasets for an analysis and easily identify what each dataset contains.

The current file display is a template on the main interface that displays the dataset currently in use. This template allows the user to easily reference what set of incidents he or she is currently analyzing. The current file display in conjunction with the new naming convention for datasets helps analysts identify what data is being analyzed.

4 USABILITY TESTING

4.1 Goal

Assess the functionality and design of WebCAT by giving users a list of tasks to perform and judging their performance based on measurable metrics.

4.2 Overview

Ten University of Virginia engineering students participated in testing. Each test subject received a verbal orientation to WebCAT and was given five minutes to explore the system. Then the subject completed a questionnaire to provide a technical profile. Next, each subject performed a list of 17 tasks. The subject’s performance was recorded using the five metrics listed in section 4.3. Finally, a follow-up questionnaire was completed to obtain a qualitative analysis of the system.

4.3 Testing Metrics

The following metrics were used to gauge the performance of each test subject:

- Number of mouse clicks to complete
- Number of navigation errors
- Time taken to perform each task
- Inability to perform a task
- Number of WebCAT errors

4.4 Testing Results and Analysis

In general, all participants accomplished simple tasks successfully, such as logging in and out. The participants did not have any trouble in viewing time and control charts. Users were able to perform these tasks in less than 10 seconds, similar to the performance with last year’s WebCAT application.

Last year’s WebCAT application did not have a functional incident matching algorithm. This procedure tested users’ ability to successfully pick a target incident and return the crime type of the most similar resulting incident. As explained in section 3.2.3, jurisdictions can take anywhere from 1 hour to several days to complete an incident matching analysis. As the testing indicates, WebCAT al-

allows students to run an incident matching query in an average of 50 seconds.

To test the GIS function, users were asked to query the dataset of robberies occurring in Loudoun County to be plotted on the map. Then, they were asked to zoom in on a point and retrieve information for that incident. 5 out of 10 clicked on the wrong button to query the information. Also, all students zoomed in and retrieved information on landmarks instead of incidents. This happened for two reasons: (1) there wasn’t a legend to distinguish between landmarks and incident points (2) all incident data points were layered on top of each other since they all had the same latitude and longitude. Thus, a user only saw one incident point and around 50 landmark points. As stated in section 3.1.2, only a few departments have GIS and those that do cite that the software is not user friendly. 8 out of 10 students performed the correct procedure in acquiring information about the data point within an average of 62 seconds. This suggests that WebCAT’s GIS function is user friendly and requires little time to complete. It is a vast improvement over WebCAT’s previous GIS tool and other GIS mapping methods used.

5 CONCLUSIONS

5.1 Summary

The capstone group has added and improved many components to the WebCAT system. Consequently, these improvements have made WebCAT more efficient in achieving its stated goal: to serve as a tool which improves cross-jurisdictional data sharing and crime analysis at the state and local levels.

Incident matching capabilities have been implemented to identify similar crime incidents using a robust advanced search algorithm. The algorithm used in WebCAT’s incident matching tool is used to identify patterns and hidden relationships based on similarities in crime characteristics. A more powerful and robust GIS mapping system was added to improve the visual representation of criminal incident data. A new, more intuitive, user interface has been implemented to minimize the learning curve for WebCAT. Finally, the administration tools functionality has been expanded to provide a means for managing user profiles and administrative privileges in WebCAT.

5.2 Interpretation

The improvements mentioned in the preceding section allow analysts who use WebCAT to (1) more effectively share criminal data and find relationships within the data, and (2) better allocate their resources to achieve the goal of improving their crime analysis and response capabilities.

Hence this application has been judged successful in terms of (1) improved quality of analysis; (2) increased speed of performing an analysis; (3) minimal cost necessary to design, implement, use, and maintain the application.

Though the application has been judged successful, there are still many limitations, as is expected for any application early in its development. WebCAT must still be tested in an operational environment to fully assess whether it is a successful tool for crime analysts. Finally, there are security considerations that have not yet been addressed, such as the confidentiality of data, and the restriction of certain components of WebCAT based on a user’s administration status.

5.3 Recommendations

The most pressing need related to this project is the deployment of WebCAT to an operational environment, to be used by current crime analysts. The police department in Richlands, Virginia has already been confirmed as the location where WebCAT will first be deployed. The WebCAT application must be implemented and tested in this new environment. Distributing and collecting user evaluation surveys will help assess the true improvement WebCAT provides to crime analysts. As WebCAT evolves, the XML specifications should be expanded to include more data fields from NIBRS. This will allow analysts to perform a more in depth analysis. Finally, security issues need to be addressed to ensure that there is no unrestricted access to confidential data or privileged components of the WebCAT application.

REFERENCES


Virginia Crime Analysis Network (VCAN) Conference.  

AUTHOR BIOGRAPHIES

**JOSH HENDRICK** is a fourth year Systems and Information Engineering (SIE) student at the University of Virginia. He is currently undecided for his plans next year. He can be reached at <jsh4y@virginia.edu>

**THOMAS HOWELL** is a fourth year SIE student at the University of Virginia. He can be reached at <tjh2a@virginia.edu>

**DARRIN LONDON** is a fourth year SIE student at the University of Virginia. He will be working at Booz Allen & Hamilton as a Level 1 consultant starting in July 2004. He can be reached at <dml4c@virginia.edu>

**ERIC LUEHRS** is a fourth year SIE student at the University of Virginia. He is going to the United States Navy’s Nuclear Reactor School to train to become a Submariner. He can be reached at <eml3j@virginia.edu>

**MICHAEL SALIBA** is a fourth year SIE student at the University of Virginia. He will be working for Deloitte Consulting next year in their Washington D.C. office. He can be reached at <ms8eu@virginia.edu>