Computer Science and Intelligence:
Building the Security Database TOC-Search

Aleksandar Smiljković[1], *Zorica Stanimirović*[1], *Darko Trifunović*[2]

Terrorist and Organized Criminal Search Data Base (TOC-search) is an interdisciplinary joint project of the Faculty of Mathematics and the Faculty of Security Studies, University of Beograde, supported by George C. Marshal European Center for Security Studies. The TOC-search is an online search engine that contains information related to global terrorism and organized criminal. It includes the most exhaustive categorization of the stored data, compared to existing databases. The TOC-search is the first database that provides extensive information on links and relations between the terrorist and organized criminal categories. Several efficient search tools (simple, advanced, in- depth and fast multiple search) are implemented in order to enable efficient exploration of all data categories. The TOC-search database has both educational and applicative function, since it helps researchers, analysts, students and others working to prevent terrorism and organized criminal and represents additional powerful measure for protecting state borders.

1. Introduction

Terrorism and organized criminal represent one of the greatest security threats to the modern human society. The two key aspects in security intelligence are collection and adequate processing of the relevant data, which includes data classification, verification, storing, exploration and analysis of the stored data. For functionality, reliability and efficiency of the data base it is necessary to apply adequate methods for data collection, verification and analysis, while data storing, exploration and analysis assumes the usage of modern information technologies and software.

The TOC-search development team consists of a group of students from the Faculty of Mathematics, University of Belgrade, working on the development and implementation of the software and administration of the database, while students from the Faculty of Security Studies work on data collection, classification, data storage and verification.

The stuff of the George C. Marshal European Center for Security Studies – GMC [1] is providing us with the reports on recent events related to global terrorism and organized criminal. The main data source is the PTSS Daily, a special service supporting PTSS counterterrorism course [2] at the George C. Marshall European Center. The PTSS is five-week course, designed for government officials, military officers and police administrators currently working in mid- and upper-level management positions of counterterrorism organizations throughout the world. The PTSS Daily newsletter is created from open sources media reports for educational purposes only. The newsletter is produced by postgraduate students and verified by senior experts and counter terrorism officers from the GMC. This is the main reason why we have decided to use PTSS as the main source.

Besides the PTSS, the GMC keeps us sending other open source reports, such as: DHS Reports (Department of Home Land Security Reports), Terrorism Open Source Intelligence Report (TOSIR), Insurgency Literature Review (ILR) and Terrorism Literature Report (TLR). Genuine PTSS and other GMC reports dating from 2006 are stored in the database in the original PDF format, and may be viewed by all TOC-search users. The students' team from the Faculty of Security Studies is studying incoming PTSS reports, extracting and storing data in the TOC-search, but other open sources are used as well. The TOC-search database also contains confidential data, which are stored by security experts form government, research and other institutions and they are not publicly available.

The TOC-search has two main functions: practical and educational. It represents a secondary tool in preventing terrorism and organized criminal. It is currently being used by professionals and practitioners in security services in Serbia and several other countries. For example, it was already used for security operations of the 2008 Olympic Games in Beijing (before database was released online), and the 2010 World Football Championship. Another aspect of the database is educational one: it is exploited as research and educational tool by students of the Faculty of Security Studies, George C. Marshal Center and numerous experts and practitioners worldwide. On the other side, the students

working in the TOC-search team gain valuable practical experience in data collection, classification and verification methodology. For most of the students in the TOC-search team, it was the first time to work on an interdisciplinary project, which helped them to develop and improve their research and learning skills and expand the knowledge outside the frame of basic courses at the faculty.

2. TOC-search data

The data in the TOC-search database are classified in the following categories: Individual, Group, Organization, Supporter and Action. The existing or suspected relations between the items belonging to different categories, or to the same category, are stored in Links section, which is a novelty compared to existing databases. All mentioned categories are linked with the GMC report section via carefully chosen keywords. The most important information from the PTSS reports is highlighted in the form of short news. They are shown on the home page of the TOC-search as "Flash points". The TOC-search data categories are presented in the Figure 1.

On the Internet one can find numerous databases with similar content, but they do not have such detailed data categorization as the TOC-search. Most of them contain data related to terrorist incidents, groups and organizations only. The most famous and the most cited data basis *GTD (Global Terrorism Database)* [3] contains large amount of data (over 80 000 records), related to terrorist incidents: date, location of the attack, target, responsible group/organization and the number of victims. *The Worldwide Incidents Tracking System-WITS* [4] is also mostly focused on actions. *The Database of Naval Postgraduate School, Dudley* database [5] provides detailed descriptions of terrorist groups only.

One of the advantages of the TOC-search database is the category "Links", which contains information on certain or possible relations between the items of the same or different data categories. This is a unique feature of the TOC-s which no other terrorist database has up to now. The stored links enable us to develop a mathematical model of terrorist/organized criminal network and to apply adequate mathematical methods for prediction of terrorist threads over time in particular region, preventing the expanding of this network or disabling the communication within the network.

Another advantage of the TOC-search database is the dynamic of the stored data. Once stored, the data does not remain static; changes and updates are possible in any moment, which is done by TOC-search moderators. Four
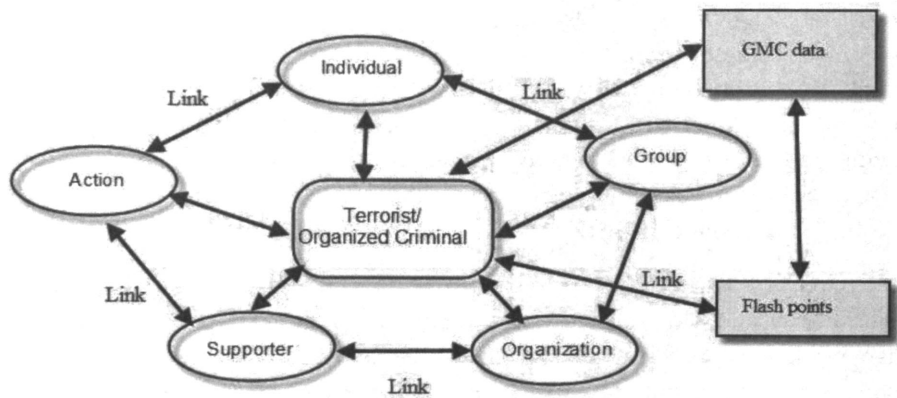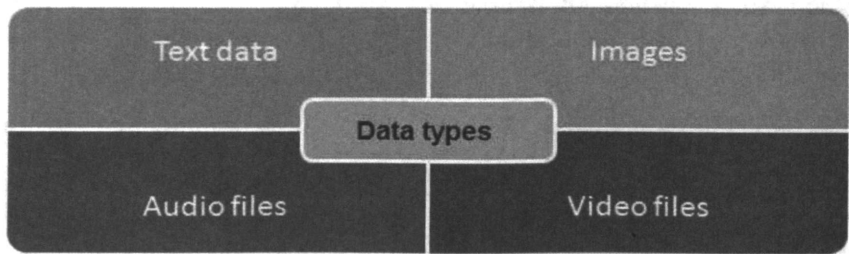
Figure 1: TOC-search data categories.

Figure 2: TOC-search data types.

data types are supported in the database: text data (various formats supported), images, video and audio files (see Figure 2).

In order to provide the protection of confident data stored in the database, two levels of data access are implemented in the TOC-s. The first level is named "blue key" and it is available for students and researches in academic institutions and research centers. The "blue key" enables the access to all open-source data stored in the base, such as: GMC reports, studies, research works, manuscripts and electronic articles related to terrorism and organized criminal that are publicly available. The second level of data access, named "red key" is reserved for legal authorities, state institutions and state government. The "red key" opens the part of the TOC-s with confident data which are valuable for all legal

institutions that are dealing with counterterrorism. The owner of the "red key" also has all rights and access to the open source data, as the "blue key" owner.

3. Technical description of the TOC-search.

3.1. Database software

The database is implemented by using a "client-server" architecture. All system services are available on the Internet. The users of the database need an Internet browser to access the system, and no other additional software is required. The TOC-search system is located on a centralized server, which holds both database application and TOC-search data. The user does not need to know the physical location of the database in order to use it. The applied "client-server" architecture has several advantages. The most important ones are:

- **Portability** – database users may be located anywhere and with an Internet browser they may exploit all functionalities of the database;

- **Flexibility** – since the system is constantly being improved, the updates are frequent; the system is automatically updated on the whole level requiring no action from users;

- **Platform independence** – the database works under any operative system.

For database software implementation we used open source tools and techniques only. The application is written in *PHP 5* programming language [6] and it is run on *Apache 2.2* server [7]. For database interface we used *XHTML 1.0* [8], *CSS 2.1* [9] and *Jquery 1.2.6 javascript framework* [10]. The database is implemented regarding the *MySql 5* [11] relational database model.

3.2. Security issues

Regarding the nature of the TOC-search database, the registration of users is required, which is realized in two steps. A potential user is first asked to fill in the registration form with some personal information and basic company information (optional). After the verification of these data, system administrator allows (or not) the access to TOC-search and at the same time prescribes the level of access to the stored data.

All login data is one-way encrypted with customized methods and stored in the database. One-way encryption additionally protects database users and provides higher security level of the software. A modification of the *MD5* algorithm upgraded to our needs is used for data encryption. All user data is available only to administrators of the TOC search. Personal information is not available to publicity or other *TOC search* users without permission. In order to ensure data privacy, all records marked with red key are available only to users with red key access and administrators. No data is deleted from database, only display status can be changed.

Regarding the security issues, the users of the database have different roles in the system, that assume different rights and responsibilities. The users are divided into two groups: external and internal users. The **external users** are allowed only to search and view the data, with no permission to change any of the recorded files. Depending on data type that an external user may access, they are further divided into "blue key" and "red key" external users. They also have the possibility to edit personal information, select the data of interest and to send feedback information. The **internal users** are divided into: "blue key" moderators, "red key" moderators and administrators. Compared to the external users, moderators are additionally allowed to store the new information on the system and to edit already stored data. "Blue key" moderators may store and edit the open-source data only, while "Red key" moderators may enter and change both open-source and confident data. Administrators have the same possibilities for data managing as the "red key" moderators, but they are also allowed to manage the users and the "flash points" section.

3.3. Search tools

The TOC-search features are implemented for the following data categories: individuals, groups, organization, supporters, links, actions and GMC reports. User interface is simple and compact, so that users can reach wanted information through minimal number of steps.

The **Simple search** is realized by keywords; while entering a keyword, the system analyzes the entered part of a keyword by sending an asynchronous query and offers suggestions for the keywords that match with the entered part. This option improves the efficiency and comfort of the simple search feature. The interface of the simple search tool for the category Individuals is shown on Figure 3. As a result of the simple search tool the user obtains the list of items that match search query or the message that no item is found. By selecting a

particular item, the system displays:

- **Basic information:** Name, Aliases, Locations, Birth date/Date of foundation Sex, Image.

- **Comment/Biography:** More detailed description of entity not covered by entity attributes

- **Terrorist/Organized criminal activity:** List of activities of the entity (bomb attack, suicide attack, armed attack, rocket attack, hijacking, conspiracy, ...)

- **Links:** Shows links of the entity with other categories

- **Multimedia files:** Multiple database entities can be connected with one or more multimedia items (video, audio, text)

- **GMC reports:** List of all PTSS (and other GMC reports) that contain some information on the entity

- **Access level:** Red key data (confidential) or Blue key data (open-source)

By using the **In-depth search** tool, the user may investigate presented details on the found subject (links, GMC reports, multimedia files,...). The **Selection tool** enables the user to select and save search results for further analysis. Since the simple search tool often gives large number of found items for the same keyword, it was necessary to implement the **advanced search** feature. In cases when certain information on the subject is known, it is more efficient to use the Advanced search tool which is enabled for the same database categories as Simple search. In the Advanced search form, the user simply enters information on the wanted item (name, aliases, location, status, activities, ...). An example of Advanced search form for category Individuals is shown on Figure 4.

By selecting the item "+", the Advanced search form expands and displays the list of options belonging to the particular group. The asynchronous query for the entered part of the group's/ organization's name is enabled in the "Membership" group, offering the user the list of groups/organizations stored in the database. The partly expanded Advanced search form for Individuals is shown on Figure 5. The Advanced search forms for other TOC-search data categories are similar.

Figure 3: Simple search form for category Individuals

Figure 4: Advanced search form for category Individuals

The **Fast multiple search tool** is additionally implemented for the category Individuals, which enables to enter several names in the same search form (Figure 6). It has great potential to be applied on the state borders when it is necessary to obtain prompt information whether individual(s) are stored in the database or have some relations with the stored data.

4. Conclusions

Ensuring the security of human society and fighting against different forms of threats, including terrorism and organized criminal, requires a comprehensive and interdisciplinary approach, which exploits the knowledge from
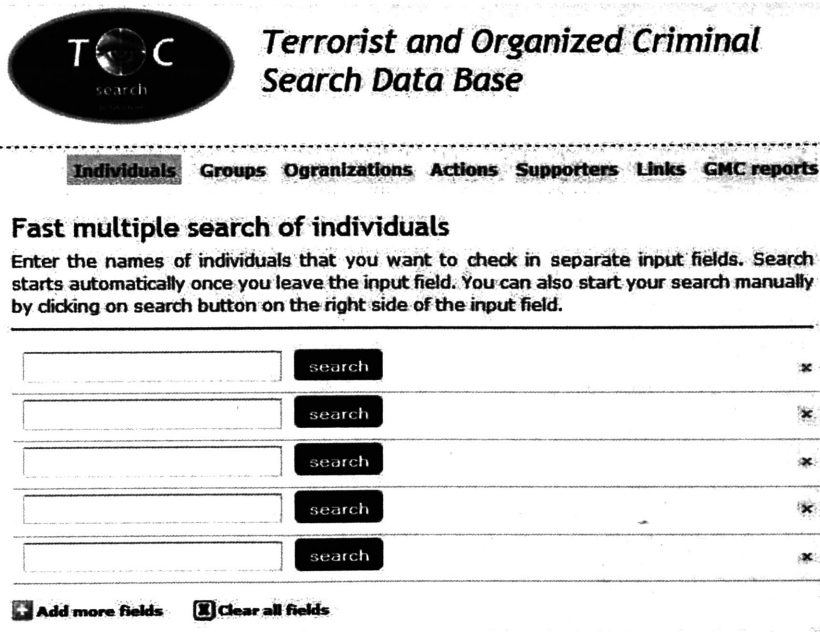
Figure 5: Partially expanded advanced search form-category Individuals

Figure 6: Fast Multiple Search Form

different research areas supported by practical experience. The TOC-search project is based exactly on these principles by interlacing computer science and mathematics with security science and intelligence.

Carefully defined categorization, links between database categories, efficient exploration tools and innovative fast multiple search feature are the main characteristics that clearly distinguish TOC-search from other publicly available databases. The TOC-s search database has great potential to be applied as an additional practical tool in public safety, state borders protection, high risk events, such as concerts, exhibitions, sport events, etc. The database also has an important educational role in tuning analytical and research skills.

In the next phase of the TOC-search project, several analytical features are to be incorporated within the database. Analytical tools will provide statistical analysis and the predictability in terms of the existence of trends, seasonality, and periodicity of terrorist and organized criminal events. Our team is intensively working on application of statistical models for a time-series analysis

of the occurrence and targets of terrorism and organized criminal events and victimization rates.

References

[1] The George C. Marshall European Center for Security Studies, http://www.marshallcenter.org.

[2] Program on Terrorism and Security Studies PTSS http://www.marshallcenter.org/mcpublicweb/en/nav-college/nav-academics-resident-courses/nav-col-ptss.html

[3] Global Terrorism Database (GTD), www.start.umd.edu/gtd/

[4] WITS, Worldwide Incidents Tracking System https://wits.nctc.gov/

[5] Naval Postgraduate Scholl Dudley Knox Library http://www.nps.edu/-Library/Research/SubjectGuides/SpecialTopics/TerroristProfile/-TerroristGroupProfiles.html

[6] PHP: Hypertext Preprocessor, http://php.net/

[7] Apache HTTP Server Project, http://httpd.apache.org/

[8] XHTML^{TM} 1.0 The Extensible Hyper Text Markup Language http://www.w3.org/TR/xhtml1/

[9] Cascading Style Sheets Level 2 Revision 1 (CSS 2.1) Specification http://www.w3.org/TR/CSS2/

[10] Jquery, JAVA Script Library http://jquery.com/

[11] MySQL-Open Source Databasehttp://www.mysql.com/

[1] *Faculty of Mathematics*
University of Belgrade
Studentski trg 16/IV
11 000 Belgrade, SERBIA
aleksandarsmiljkovic@gmail.com
zoricast@matf.bg.ac.rs

[2] *Faculty for Security Studies*
University of Belgrade
Gospodara Vučića 50
11 000 Belgrade, SERBIA
galileja @yahoo.com