COMPUTER DATA SEARCH AND COMPARISON - GENERAL REVIEWS AND APPLICATION IN CRIME INVESTIGATION

[*]Marinkovic D.[*], Brankovic A.[1], Milojkovic B.[1]

Criminal Justice and Police Academy, Belgrade, Serbia

Abstract: Collecting the most versatile kind of information about the citizens and their storing in the appropriate bases represent the reality of the contemporary society. The growth in the quantity of these pieces of information has exceeded human power to process and analyze such huge quantities of data in a traditional manner, requiring computerized techniques and means for these needs. Although widely applied for years in the work of public administration and economy, so far the computer data search and comparison have not been sufficiently used in crime investigation and forensics. Police agencies and forensic laboratories collect large quantities of various data which originate as a result of processing numerous criminal activities. The very success of their automatic search and comparison within criminal investigations depends to a large extent on the availability and characteristics of data (features, raster) which refer to persons, objects or events.

Key words: computer data search and comparison, data mining, computer matching, data surveillance, criminal-investigation aspects of data search and comparison, forensic data bases

1. Introductory remarks

Exceptional organization of human society which is present today not only in the developed countries but in other, in many ways underdeveloped parts of the world, involves collecting and managing the most various data related to their members. The efficient functioning of the government machinery and non-governmental sector requires the existence of many information registers about individuals and legal entities, their lives and work related to a specific field or problems because of which such registers and data bases are kept. On the other hand, the very development of computer technology (computerization) has largely

[*]Corresponding author: e-mail: darkoart2003@yahoo.com

increased the possibility of receiving, processing and monitoring such data, even for the purpose of surveillance of individuals and their behaviour. The essential importance of computer processing and storing of information is not only in the speed of carrying out various operations, but primarily in the possibility to access the integrated mutually linked elementary data which come from different sources. The state-of-the art information technology makes it possible to get these data in the matter of seconds or parts of a second, by networking data bases within large state and social areas, such as public administration, economy or science.

Collecting relevant information on citizens out of various (naturally legal and legitimate) motives and for the most various purposes as well as their storing into the appropriate bases represents the reality of the contemporary society, in the same way as it is the realistic (and necessary) fact that the persons these data refer to cannot have the absolute power over them. However, they have the right to feel secure from possible misuses of these data. This is why the issue of the protection of citizens' data today is highlighted even more, particularly being prominent in functioning and performing the activities of state administration institutions and judiciary, including police. Accordingly, with regards to the availability of citizen-related data, they must have certain limitations for the general interest in the same or similar way as when their other freedoms and rights are limited. The task of the legal science, law-makers and legal practice is to define standard foundations for the collection and management of the most versatile data, i.e. the conditions under which they can be used for socially justified purposes. On the other hand, the actual (primarily technical) possibilities are increasing from day to day for more comprehensive, complex and sophisticated exploiting of data on man and his activities in all fields of life and work. Among other things, the exploiting of such data can yield good results in fighting crime as well.

The explosive growth of quantity of data and bases where they are stored has exceeded human power to process and analyze such huge quantities of data by traditional means, requiring new and different techniques and means of automatic analysis in the available bases. Automatic data search and comparison, regardless of the purpose they are used for, is based, on the one hand, on the bases where certain data are stored, and on the other hand, on the application of computers (understood as hardware) and related programs (software) used for the search, comparison analysis of these data.

2. Data surveillance as a special form of personal surveillance

Surveillance may be defined as a systematic investigation or monitoring of movements or communications of one or more persons in order to collect information on them, their activities and connections. For a long time the surveillance has been implemented by direct physical observation, as well as by various devices used for the support, including telescopes, cameras, directed

microphones, telephone bugs, etc. The conventional forms of surveillance require hard work, cost much and last long (Marinkovic, 2008).

In the course of the 20th century the work of public administration has increasingly included the intensive use of personal data. The expansion of network traffic and flow of information has additionally contributed to the huge amounts of data interchanged to be widely available. Personal surveillance through personal data has become easily achievable, and at the same time much more inexpensive and simpler than the conventional techniques of physical or electronic surveillance. As a result, *data surveillance* has started developing. This is a method of surveillance of a large number of people by comparing and pairing of data referring to them which have been collected from a large number of sources. Ever since it started being applied, the data surveillance has become a topic of numerous government publications and its effects and influences have been discussed by many sociologists and some lawyers. It is usual that in Anglo-Saxon literature this phenomenon is referred to as *dataveillance* while it essentially represents the control, comparison and analysis of systemized data on persons in investigations or monitoring of their activities. There are two essential models of personal surveillance through data, and these are: 1) *personal dataveillance*, such as checking or validation of concrete, extraordinary or extra works and transaction, which are contrary to internal regulations of a certain service or organization, and 2) the surveillance of a large, usually unidentified number of persons (*mass dataveillance*), such as checking and validation of all transactions which are contrary to internal regulations of a certain organization. In addition to the two above mentioned models, there are also *facilitative and support techniques*, such as techniques for integration of data stored in various databases. In comparison to conventional forms of surveillance, dataveillance is automated, and therefore cheaper and more reliable. This is why its application during the last 30 years has flourished, in the beginning in wealthy societies with the developed and sophisticated information technologies, but recently also in the developing countries, among which there is a significant number of them having legislative problems due to the insufficiently developed mechanisms of civil rights protection.

3. Various models of computer data search and comparison

We are of the opinion that there should be a terminological difference between the concepts of (computer) data search and comparison. The search includes reviewing and analysis of data contained in certain data bases in order to find information that are not visible at first sight and refer to a certain person, action or process. Defined in such a way, the computer data search is mostly contained in *data mining* techniques. On the other hand, comparison implies to

have a certain amount of data or features in advance, which is then entered and compared with other data from a certain data base in order to find common characteristics between them which connect them and make them similar or the same (pairing). The procedure of computer comparison is almost entirely made equal with the procedure of *computer matching*.

In various fields of research (primarily statistics and artificial intelligence) the procedures of automated analysis have been developed that reveal hidden contents within large sets of data. The process used to achieve this is usually called *data mining*.[†] It marks the automated analytic process shaped for the effective and efficient exploration in large data sets in order to reveal and use valuable, "hidden" information which refer to hitherto unknown facts and relations. In other words, *data mining* can be understood as finding the previously unknown and potentially useful information or knowledge from large data sets. The basic principle is to create computer programs which scan such data sets and automatically search for certain, previously defined patterns. The potential of *data mining* technologies depends much on the nature of the available data sets and it is successfully applied in various professional fields, for instance in the remote resource management, biometrics, speech recognition or business and marketing. The *data mining* procedure uses algorithms in order to find the important hidden contents in large sets, the interpretation and understanding of which enables better diagnostics of state of affairs, better predictions and finally better decision-making.

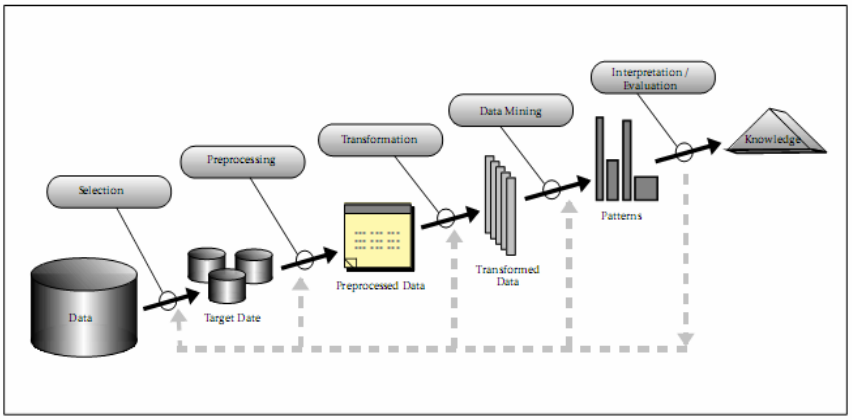

Figure 1 - Knowledge Discovery in Databases – KDD, in which Data Mining makes an integral part of the process (Fayyad et al., 1996)

[†] This is a metaphoric term used to present this process comparing it with ore mining. In the same way as ore mining is rather difficult and uncertain job, when you search for a certain precious ore in the depths of the Earth, during this procedure you dig, in other words search the vast amount of data looking for those that are useful.

The basic functions of *data mining* are: 1) classification, i.e. exploring of entity features and their sorting into previously determined classes; 2) clustering, i.e. segmenting of a heterogeneous set of entities into homogeneous sub-groups, clusters; 3) evaluation, i.e. predicting of unknown values of continuous variables; 4) detection of changes and deviations of data from previously measured or standard values; 5) detecting associations and finding items in transaction which imply the presence of other items in the same transaction, etc. Some authors (Berry, Linoff, 2000) classify *data mining* functions into two sets – the first one is directed analysis, based on supervised learning, including classification, evaluation and prediction, and the second one is undirected analysis, based on unsupervised learning, including grouping, association rules, description and visualization. The dominant view of the nature of *data mining* is that it helps reveal just the hypotheses of complex facts and their relations (Fayyad et al., 1996).

One of the *mass surveillance* techniques is computer matching, i.e. data pairing, which includes matching of (computer automated) readable records which contain personal data (generalities) of a large number of persons in order to reveal and clarify interesting cases. This technique is called *computer matching* in the USA, or *data matching* in Australia and Canada. It has become economically feasible in early 1970s, as a result of the information technologies development, and it has been developing since then so that nowadays it is widely applied, particularly in the sphere of state administration of the three mentioned countries. Some of the forerunners of *computer matching* could be found in so-called *Income Matching Programs*, which were long used by the USA IRS, or by the system for parental help, approved by the USA Congress by the amendment on the *Social Security Act* in 1974. Its original goal was to find parents who have violated the agreements related to their child support and to make them competent to honour and implement such agreements (Clarke, 1994).[‡]

[‡] It is stated in the literature that the first computer program intended for comparing and matching of data was the so-called Project Match, implemented in 1977 in the USA by the then Department of Health, Education and Welfare. Project Match compared the data of approximately 78% of the total number of families that received child support, with the data from salary lists of about 3 million federal officers. 33000 raw matches were reported, the number then being reduced to 7100, which resulted in 638 cases of internal investigations and 55 charges.
It is estimated that until 1982 about 200 programs for organizing and comparing of data were routinely carried out by the state and federal agencies in the USA. President Reagan's administration launched an action to improve government efficiency, and the President's Council on Integrity and Efficiency in Government (PCIE) has become the most ferocious advocate of computer matching method introduction into the contemporary management. Congress' Office of Technology Assessment estimated that the number of applications of computer matching method

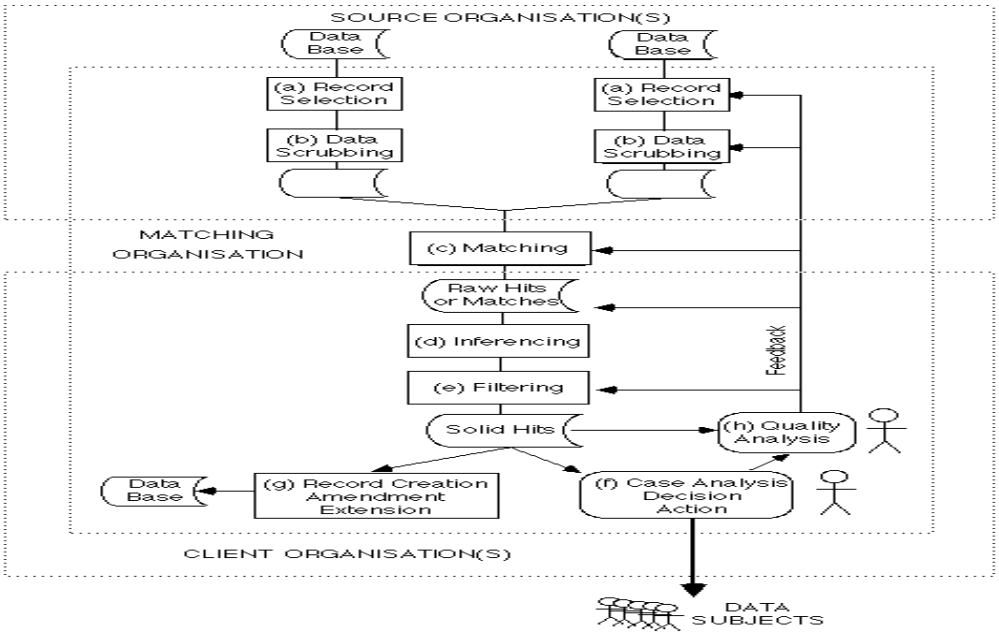

Figure 2 - Conceptual model of computer matching process (Clarke, 1994)

Computer matching technique is used for various purposes, the majority of which refer to social control and efficient work of state administration (traffic, police, health insurance, and similar), while its goals can generally be divided into primary and secondary ones. Some of primary goals would be: 1) to discover errors in programs of administration institutions (for instance, faulty estimate of certain profit, issuing bills several times, etc.); 2) to verify whether the conditions required for further use of certain benefits have been fulfilled; 3) to discover unlawful conduct of tax-payers, users of certain benefits, government officials, and similar (false or multiple claims, undeclared income or property, inappropriate conduct, conflict of interests); 4) monitoring regularity of procedure for allocation of grants or concluding agreements); 5) finding the addresses of persons that the government agencies have certain claims from; 6) identification of those who have the right to certain profit but do not exercise this right at the moment; 7) control of data validity, and 8) updating the data stored in one set of records based on the data from another database. Among the secondary goals of *computer matching* application, we would point out the following: 1) support to actions with favourable financial effects, such as ceasing cooperation with irregular payers, reduction of excessive payoffs, allowances for false payments to agencies, unpaid taxes or arrears due, collecting allowances in favour of other

in the period from 1980 to 1984 increased three times, while Laudon pointed out that the number was 500 by 1986. Quoted according to Clarke R., 1994.

government agencies, avoiding future irregular or excessive payoffs, intimidating and adverting future unlawful behaviour; and 2) establishing and maintaining the databases for the purpose of social control, research and statistics, improvement of strategic programs, and procedures and control mechanisms.

In addition to *computer matching*, there are other closely related techniques used for the support of surveillance of wide layers of population using data. One of them is *data-linkage*, the purpose of which is to store individual records into one personal file through which it is possible to identify one or more other files, which enable fast and reliable interrelation between the data in the future. The second technique, known as *data concentration* includes linking and joining of databases and creation of new ones for the requirements of support to numerous functions of state administration and economic subjects. The third technique includes the use of *common multipurpose identifiers*, which has aroused many debates on the creation of wide national programs intended for the personal identification, such as bases with social security numbers in the USA and Canada.

It is not rare that a person wants to acquire a certain profit in a fraudulent manner, for instance to receive a bigger pension cheque by giving false data about his family condition or to pay lower taxes because the tax authorities do not have true knowledge about his income; or get a loan although he does not fulfill the conditions because the creditor is not acquainted with the fact that the loaner has already the outstanding debts which are due. Under such circumstances, the organizations will probably require the confirmation of accuracy and completeness of the data enclosed by the interested persons. In order to protect their interests, they carry out the *verification* procedure, in other words they check if the presented data are true. The term *verification* is used as a common term for these purposes, but considering that it implies higher standards of proving and accuracy than it is possible to determine in these cases (without the court procedure), the term *cross-checking* is naturally more appropriate.

The large part of processing of and manipulating data is internal and is carried out for the needs of one organization. However, cross-checking in general implies the use, or discovering and disclosing of data in specific cases, which were collected earlier and processed for other functions and/or within other organizations. *Cross-checking* can be carried out in *ad hoc* situations, as required, or according to previously regulated agreements between certain organizations. They can be carried out with or without the knowledge and/or consent of the individual, as well as with or without an explicit legal authorization. Many *cross-checking* activities start on the occasions when certain individuals apply for a job, pension or loan, in which case it is usual to call them *front-end verification*. Reverse or inverse arrangement includes the agreement between organizations, which implies automatic cross-informing in case there are changes of certain data, for instance of the address. Such a procedure could be marked as *front-end*

notification. *Front-end verification* and *front-end notification* are the models of data surveillance, as a set of techniques by which one or many individuals are controlled but not directly, by physical surveillance, but using data. The previously mentioned cases, where the monitoring is actually a specific identification of persons resulting from the transactions which include the data related to these persons represent the forms of *personal dataveillance*. The person subject to such a manner of surveillance can be marked as *digital persona*.

Cross-checking can be undertaken even without the initiative by the subject who should carry out the transaction related to a certain person. The reasons for this may be contained in lifting the suspicion of the honesty of a client and the belief in their inclination to frauds, as well as validation of the data related to persons with whom the organization cooperates in order to avoid potential damaging consequences. In addition to assistance in the implementation of *personal dataveillance*, *cross-checking* may give large support to the implementation of *mass dataveillance*, which may be undertaken because it is not possible to identify in advance those persons who can be put in the category of the suspicious ones, or those inclined to embezzlement.

4. Crime-investigating aspects of computer data search and comparison

Computer search, analysis and comparison of data for crime-investigation purposes may be versatile, with various expectations and results of application. In the same manner as the large number of data stored in appropriate databases serves to the efficient performance of public administration, administration or banking, it can be very useful in crime investigation. From the point of view of the crime suppression activities, databases can be divided into primary and secondary. Primary databases are those created and maintained primarily for the requirements of crime investigations and subjects working on them, while the secondary ones are those organized and managed for the requirements of state administration, economy or health, but in certain cases they can be used for crime investigation purposes. This means that primary databases include, for instance, fingerprints data bases or criminal DNA profile databases,[§] while the secondary

[§] The Interpol's automatic fingerprint database (AFIS) contains about 90,000 fingerprints belonging to offenders, as well as fingerprints lifted from 1,600 crime scenes. DNA databases contain DNA profiles which are classified into reference profiles (the profiles of offenders, victims, the aggrieved...) and trace profiles (profiles obtained from biological traces). According to Interpol's data, in 2008 the forensic DNA analysis was carried out in the majority of Interpol member-countries, 53 countries have DNA database, while it is being created in 29 countries. There is a consensus at the level of the organization that each country should have a DNA database, and that the international exchange of DNA profiles should be carried out. There are discussions about whose DNA profiles should be stored in databases. Legal regulations in the

would include the databases of money transactions carried out by certain banks or bases of tax payers.

The factors helping in the evaluation of relevance of data mining techniques application in crime suppression range from the activities from which the databases result to their quality (the degree of insecurity, precision and completeness). Police agencies and forensic laboratories collect large quantities of various data, which result from the processing of many criminal activities. Thus the group of data is obtained within the forensic crime scene investigation which consists of the information referring to collected material of physical origin (for instance, biological traces, traces of tools, fingerprints, shoeprints, and illegal drugs seizures). This kind of data may be presented numerically and may be subject to categorization. The features extracted from these materials are often imprecise (in principle because of the instruments used for the analysis and measurements), incomplete (fragmentary) and insecure.

The discovered and processed material samples are usually categorized into three groups: 1) *useless samples* (for instance, the obvious clarity of the contents without any calculations or they are irrelevant for the problem observed), 2) *useful samples*, which provide direct important information that can be worked with, and 3) *patterns that require interpretation*, and which can be classified into two previous categories, because of which they must be studied by the experts in the given field (Terrettaz-Zufferey et al., 2006).

The researchers have developed various automated data mining techniques for the requirements of crime suppression, both in the field of local police work and at the national level. Thus the *entity extraction technique* identifies the patterns from databases such as texts, images or audio materials. It is used for the automatic identification of faces, addresses, vehicles or personal characteristics from narrative police reports. This technique provides for the basic data for crime analysis, but its achievements depend to a large extent on the availability of large quantity of pure input data. *Cluster techniques* systematize data into groups of similar characteristics, in order to maximize or minimize the similarity of data within a certain group – for instance, for the identification of suspects who commit crimes in the similar manner or to differentiate between criminal groups belonging to different gangs. *Association rule discovery* finds the groups of data that appear often in one database and the patterns of their appearance are defined as regularities. This technique is often used to trace computer network hacking so that the certain rules of association could be deduced from the history of interactions among the users. The researchers can also use this technique for

Interpol member-countries are various, ranging from Belgium where the DNA databases contain only the profiles of those convicted for major crimes, to Great Britain where the databases contain the profiles of both the suspects and convicts for all crimes and the majority of delicts, as well as the profiles of volunteers. Quoted according to INTERPOL – Forensic.

profiling of hackers so that they could help in detecting possible attacks on the network.

Sequence pattern detection (or string pattern detection) finds sequences that appear often in one set of transactions that occurred at various times. Pointing to the hidden patterns is useful for crime analysis, but in order to obtain meaningful results a rich and highly structured database is required. *Deviation detection* uses certain measures for the study of data which noticeably differ from other data. The investigators may use this technique to detect frauds, hacking into network systems and other crime analyses. However, such activities may sometimes seem usual at first sight, which makes identification of deviating data more difficult. *Classification* finds common features among various criminal entities and organizes them into previously defined classes. This technique is used for the identification of so called spam e-mail messages, based on the linguistic patterns and structural features of the sender. Often used for prediction of crime trends, the classification may reduce time required for the identification of criminal entities.

Comparative data mining techniques compare pairs of textual fields in databases and calculate similarities between records. These techniques may discover false information such as names, addresses and social security numbers. The investigators may use comparison for the analysis of textual data, but these techniques often require intensive calculations. *Social network analysis* describes the role and interactions among branching points (nods) within one conceptual network. This technique may be used in case the networks were created which would illustrate the roles of certain criminals, the flow of material and immaterial goods and information, as well as connections between these entities. Further analysis may reveal critical roles and sub-groups, as well as vulnerability, i.e. weaknesses within the network (Chen et al., 2004).[**]

One of the aspects of applying *data mining* techniques for crime-investigating purposes is the analysis of seized drugs in order to define the status of drug market as complete as possible (Ratle et al., 2006). In this case the methods of recognizing drug samples are systematically tested on the multitude of samples of seized heroine and cocaine in order to find possible regularities which could provide information related to the scope and development of illegal trafficking. Classic algorithms, such as the analysis of main components and various grouping and classifying algorithms, can successfully be applied on heroine databases. Basically, the process of diluting and cutting heroine happens at various levels of illegal trafficking, but it is most often carried out at the end of

[**] The offenders often develop criminal associations – networks within which they make groups or teams in order to commit various illegal activities. The application of data mining techniques in these cases consists of identification of sub-groups and key members in these networks, as well as of the study of patterns of interaction in order to develop efficacious strategies for neutralizing of these networks. More in Chen H. et al., 2004.

the distribution process so that the quantity of pure heroine is as small as possible, in other words the profit is as large as possible. This is why the substances for cutting heroine are of special importance for easier understanding of local trafficking network. The presence or absence of these substances is systematically detected by laboratory techniques of chemical analysis. One sample of the seized heroine can contain various substances at the same time (sugar, milk, pudding, or cocoa powder, flour, paracetamol and similar), and a certain combination of the ingredients and their ratio can be the indicator of various levels in the chain of distribution. This is why the dynamics of appearance of these combinations may be a good indicator of the condition and development of a local market, with the possibility of presenting by means of combining analysis and graph theory. Databases created for these purposes should contain the following variables (Terrettaz-Zufferey et al., 2006):

- location and time of seizure;
- presence/absence of cutting substances;
- combination of cutting substances.

When we talk about the application of computer data matching methods in crime investigation, the starting basis is made of the available features of a certain person or things or kind of events (criminal act, misdemeanor, the procedure of determining ownership, and similar) because of which the matching is carried out in the first place. It is based on these that databases are determined where it is expected to find complementary data referring to that person(s), things or events. Personal features or characteristics may be related to his personality, taken in psycho-physical (sex, age, fingerprint, DNA profile, diseases, etc.) or social sense (nationality, citizenship, political orientation, bank account, marital status, membership in some organization, etc.). Also, the features can be such as to be characteristic for only one person, so that when they are matched the identity of the person is determined beyond any doubt, or they can be common for a big or small group of people, which are then, following the search and matching, selected from the database and processed further. Therefore, there are two kinds of computer data matching:

1. comparing the data the result of which is to determine the identity of a person (for instance, by running the DNA blood samples from crime scenes through the criminal DNA profile database or by running the dead John Doe's fingerprints through the database of identity cards of the citizens);
2. searching the data the result of which is to determine the circle, or a group of people (for instance, by searching through the database containing the data on vehicles registered in a certain area in order to select vehicles of a

certain brand, type and colour, or the owners of such vehicles, because of a car accident).

Nowadays the police forces all over the world use the *Automatic Fingerprint Identification Systems* (AFIS). In these cases, we talk about the primary databases, considering the fact that such registers are made for crime-investigation purposes. However, such fingerprint bases or the bases containing other biometrical features which include the wide range of population are starting to be created, without any specific criterion except, for instance, the age or entry to the territory of a certain country.[††] In the first case the motive is to issue such identification documents to citizens (identity cards, passports) that would contain, among other things, some biometrical characteristics, most often the photograph of the person and his/her signature,[‡‡] and in the second it is the business, tourist or any other entry in the country that requires a certain procedure.

From the crime-investigation aspect, special significance is given to data matching in cases when there are material features available which are found at crime scenes or some other places and which are (or it is assumed that they are) connected to a crime, with the material features of that kind taken from the suspects for the comparing purposes. In this way, in case all features match, their connection is determined by quite a simple procedure, or in case they do not match, persons are eliminated as suspects.

It can be said that the success of computer data matching in criminal investigations depends crucially on the availability of characteristics (raster, features) of the persons and their features. Accordingly, if a small number of characteristics are available, less is the probability that the search will be successful. On the other hand, if the characteristics are too general, a large number of persons will result from the matching process and they should be processed further, which increases the costs of investigation to a large extent. This is why

[††] For instance the US-VISIT (*United States Visitor and Immigrant Status Indicator Technology*) program requires all the USA visitors to be photographed and their fingerprint taken prior to entering the country. These data are used not only for verifying the visitors when entering the USA, but they are connected with more than 20 other databases of the USA government. The goal is to prevent to a significant extent the entry of the wanted or dangerous persons who assume false identities to enter the country. Similar to US-VISIT, Japan uses J-VIS program. More in: *Homeland Security: Fact Sheet – Expansion of US-VISIT Procedures to Additional Travelers; United States Visitor and Immigrant Status Indicator Technology*.

[‡‡] This is the situation present in the Republic of Serbia also, following the passing of the Law on identity cards, according to which this basic identification document contains, among other things, the photograph of a person, signature and a fingerprint. In this way the Ministry of the Interior would in the course of issuing new identity cards create such a database where biometrical features of all Serbian citizens older than 16 (exceptionally some younger ones, too) will be stored.

some authors are questioning sincerely the very efficiency of this evidence procedure.[§§]

5. Conclusion

The great challenge all police and intelligence agencies are facing is an accurate and efficient analysis of the data on crime, the scope of which is constantly increasing. For instance, complex criminal conspiracies are often hard to reveal because the information on suspects may be geographically scattered and may include large number of people. Disclosing computer crimes can also be difficult because the extensive network traffic and frequent *online* transactions create a huge quantity of data out of which only a small portion refers to illegal actions. Police agencies and forensic laboratories collect large quantities of various data, which result from processing many criminal activities. It can be said that the automatic data searching and matching techniques have been insufficiently used so far in this field, although it could contribute significantly, particularly in discovering these crimes which are a part of dark numbers or are difficult to anticipate and prevent. Extenuating circumstance in their application is, among other things, huge versatility of data that should be processed and considered.

Those involved in criminal investigations who have years of experience can often precisely analyze crime trends, but since the frequency and complexity of criminal acts increases, human errors also appear, the time required for analysis increases as well, and the offenders have more time to destroy evidence and avoid being arrested. Automatic data searching and matching is a powerful tool which enables the crime investigators, who may not be skilled for analysts, the fast and efficient searching of large databases. Computers can process thousands of instructions in just a few seconds, saving time. In addition to this, installing and using of software often costs less than hiring or training of the staff. Computers are also less prone to errors than people, especially those investigators who work many hours both at day and night.

Special understanding of the relationship between the possibilities of the analysis and the characteristics of a certain type of crime can help investigators to apply these techniques more efficiently in order to identify trends and patterns, locate problem area, and even predict a crime.

[§§] In Germany in 2004 the fact was made public that the search of as many as 8.3 million data resulted in only one investigation, which strongly supported the arguments of critics that the search raster is actually a pure failure. Quoted according to *Rasterfahndung – Kritik*.

References:

1. Berry M., Linoff G.,(2000.), *Mastering Data Mining*, New York

2. Clarke R., (1994), Dataveillance By Governments: The Technique of Computer Matching.- In: *Information Technology & People*, p. 46-85**;** http://www.rogerclarke.com/DV/MatchIntro.html (downloaded on June 10, 2009).

3. Chen H. et al., (2004), Crime Data Mining: A General Framework and Some Examples.- In: *Computer*, Published by the IEEE Computer Society, p. 50-56.

4. Fayyad U. M. et al., (1996), From Data Mining to Knowledge Discovery: An Overview.- In: *Advances in Knowledge Discovery and Data Mining*, Cambridge, p. 1-34. http://www.daedalus.es/fileadmin/daedalus/doc/MineriaDeDatos/fayyad96.pdf (downloaded on August 18, 2009).

5. Homeland Security: Fact Sheet - Expansion of US-VISIT Procedures to Additional Travelers; http://www.dhs.gov/files/programs/gc_1231972592442.shtm (downloaded on April 24, 2009).

6. INTERPOL – Forensic; http://www.interpol.int/Public/Forensic (downloaded on June 28, 2009).

7. Marinković D., (2008), Tajni audio nadzor kao dokazna radnja – različiti modaliteti i analiza rešenja u zakonodavstvu Srbije.- In: *Sprečavanje i suzbijanje savremenih oblika kriminaliteta III*, Beograd, p. 228-256.

8. Ratle F. et al.,(2006), Learning Manifolds in Forensic Data.- U: *Lecture Notes in Computer Science*, Heidelberg, p. 894-903; http://resources.metapress.com/pdf-preview.axd?code=fkgv59020149w601&size=largest (downloaded on July 4, 2009).

9. Rasterfahndung - Kritik; http://de.wikipedia.org/wiki/Rasterfahndung (downloaded on June 18, 2009).

10. Terrettaz-Zufferey A. L. et al., (2006), Assesment of Data Mining Methods for Forensic Case Data Analysis.- U: *Varstvoslovje*, Fakulteta za varnostne vede, Ljubljana, p. 350-354.

11. United States Visitor and Immigrant Status Indicator Technology;

http://en.wikipedia.org/wiki/United_States_Visitor_and_Immigrant_Status_Indicator_Technology (downloaded on April 24, 2009).

REZIME

Prikupljanje odgovarajućih informacija o građanima iz najrazličitijih (naravno legalnih i legitimnih) motiva i u najrazličitije svrhe, te njihovo smeštanje u odgovarajuće baze, predstavlja realnost savremenog društva. Razvoj računarske tehnologije u velikoj meri je povećao mogućnosti prijema, obrade i praćenja takvih podataka, pa čak i u svrhe nadzora nad pojedincem i njegovim ponašanjem. Automatsko pretraživanje i upoređivanje podataka, nezavisno od toga u koje se svrhe primenjuje, zasniva se sa jedne strane na bazama u kojima su smešteni određeni podaci, i, sa druge strane, primeni računara (shvaćenog kao hardver) i odgovarajućih programa (softver) kojima se ti podaci pretražuju, upoređuju i analiziraju.

Kompjutersko pretraživanje, analiziranje i upoređivanje podataka u kriminalističke svrhe može biti veoma raznovrsno, sa različitim očekivanjima i rezultatima primene. Policijske agencije i forenzičke laboratorije sakupljaju velike količine različitih podatka, koji nastaju kao rezultat obrade brojnih kriminalnih aktivnosti.

Veliki izazov sa kojim se suočavaju sve policijske i obaveštajne agencije jeste tačno i efikasno analiziranje podataka o kriminalu, čiji se obim neprestano povećava. Može se reći da su tehnike automatskog pretraživanja i upoređivanja podataka do sada nedovoljno eksploatisane u ovoj oblasti, iako bi mogle dati značajan doprinos.

Automatsko pretraživanje i upoređivanje podataka je moćna alatka koja istražiteljima krivičnih dela omogućava brzo i efikasno pretraživanje velikih baza podataka. Posebno razumevanje odnosa između mogućnosti analize i karakteristika određene vrste krivičnog dela može da pomogne istražiteljima da efikasnije primene ove tehnike kako bi identifikovali trendove i obrasce, locirali problematična područja, pa čak i predvideli krivično delo.

SUMMARY

Collecting relevant information on citizens out of various (naturally legal and legitimate) motives and for the most various purposes as well as their storing into the appropriate bases represents the reality of the contemporary society. The development of computer technology) has largely increased the possibility of receiving, processing and monitoring such data, even for the purpose of surveillance of individuals and their behaviour. Automatic data search and

comparison, regardless of the purpose they are used for, is based, on the one hand, on the bases where certain data are stored, and on the other hand, on the application of computers (understood as hardware) and related programs (software) used for the search, comparison and analysis of these data.

Computer search, analysis and comparison of data for crime-investigation purposes may be versatile, with various expectations and results of application. Police agencies and forensic laboratories collect large quantities of various data, which result from the processing of many criminal activities.

The great challenge all police and intelligence agencies are facing is an accurate and efficient analysis of the data on crime, the scope of which is constantly increasing. It can be said that the automatic data searching and matching techniques have been insufficiently used so far in this field, although their contribution could be significant. Automatic data searching and matching is a powerful tool which enables the crime investigators the fast and efficient searching of large databases. Special understanding of the relationship between the possibilities of the analysis and the characteristics of a certain type of crime can help investigators to apply these techniques more efficiently in order to identify trends and patterns, locate problem area, and even predict a crime.