مجلة جامعة تشرين للبحوث والدراسات العلمية – سلسلة العلوم الأساسية المجلد (34) العدد (2) 2012
Tishreen University Journal for Research and Scientific Studies - Basic Sciences Series Vol. (34) No. (2) 2012

Initial Development of Technique for Early Predication of Terroristic and Criminal Behavior using XML-Technologies and Data Mining Strategies (TEPTCB)

Dr. Ammar Suhail Balouch *

(Received 16 / 1 / 2012. Accepted 29 / 4 /2012)

□ ABSTRACT □

The reports used in education's field to identify and assess the behavior of students, and reports used in the security field to pursue criminals or terrorists in the community, lack some of the criteria that help the conversion of these reports to information that can be used before the terrorists commit the criminal event. This is called "early to predication of terrorist and criminal behavior" of persons.

This early predication of mentioned behavior is the main idea of our work, because there are projects to find the terrorists or criminals after doing the criminal events. These projects identify and analyze the data from the occurred even, including some information about the area and the persons. This is the major difference between these projects, and our work in this article.

Keyword: XML, XQuery, DTD, Data Mining, Database, decision support system, crime, terrorist.

* **Assitant Professor, faculty of computer science, University of Tishreen, Lattakia, Syria.**
Competence: Database's and Information's systems, email: a.balouch@gmx.de

مجلة جامعة تشرين للبحوث والدراسات العلمية – سلسلة العلوم الأساسية المجلد (34) العدد (2) 2012
Tishreen University Journal for Research and Scientific Studies - Basic Sciences Series Vol. (34) No. (2) 2012

تطوير أولي لتقنية التنبؤ المبكر للسلوك الاجرامي والارهابي باستخدام تقنيات الـ XML واستراتيجيات التنقيب عن البيانات (TEPTCB)

الدكتور عمّار سهيل بالوش*

(تاريخ الإيداع 16 / 1 / 2012. قُبِل للنشر في 29 / 4 /2012)

□ ملخّص □

إن التقارير المستخدمة في المجال التربوي لتحديد وتقييم سلوك الطلاب والتقارير المستخدمة في المجال الاجتماعي الأمني لملاحقة الأفراد الاجراميين في المجتمع تفتقد إلى بعض المعايير التي تساعد وتساهم في تحويل هذه التقارير إلى بيانات يمكن الاستفادة منها ليس فقط بعد حدوث الحدث الاجرامي او الارهابي بل قبل حدوثه وهذا ما ندعوه التنبؤ المبكر عن السلوك الاجرامي والارهابي للأشخاص.

هذا التنبؤ المبكر عن السلوك المذكور هو فكرة العمل الرئيسية لأنه توجد أعمال في هذا المجال ولكن لإيجاد الافراد الذين قاموا بعمل اجرامي وذلك بتحليل البيانات الناتجة عن العمل الإجرامي وهذا هو الفرق الرئيسي بين هذه الاعمال والعمل الذي قمنا به والمراد ايضاح فكرته في هذا المقال.

الكلمات المفتاحية: XML، XQuery، DTD، تنقيب البيانات، أنظمة دعم القرار، الجريمة، الارهاب.

*مدرس – قسم هندسة الذكاء الصنعي – كلية الهندسة المعلوماتية – جامعة تشرين – اللاذقية – سورية

Introduction:

The benefit of data mining and XML for Society, parents, and country security seems tremendous. Yet only a few limited applications are documented. This paper introduces a new approach that tries to overcome these problems in the form of Data Mining systems with associative memory as the main technique.

The reports used in education's field to identify and assess the behavior of students and reports used in the security's field to pursue criminals or terrorists in the community lack some of the criteria that help and contribute to the conversion of these reports to information can be used before occurring of the terroristic or criminal event, and not after it occurs. This is called "early to predication of terrorist and criminal behavior" of persons.

This early predication of mentioned behavior is the main idea of our work, because there are projects to find the terrorists or criminals after doing the criminal events. These projects identify and analyze the data from the occurred event then it can find some information about the area and the persons. This is the major difference between these projects (works) and our work in this article.

In this work we have used the modern technologies of Data Analysis and storage of the data in format of XML-documents. For getting the early predication of terroristic and criminal behavior we have used database system to build a data-warehouse for our project using defining data mining rules. These points are supported with the query language for XML- documents called XQuery.

The proposal of this work is expected to be the basis for building a decision support system for the early prediction of criminal and terroristic behavior.

Related Works:

Related works can be summarized in the following:

The works [1], [5], [7], [9], [17] analyze the crime and draw on the crime map, while the works [8], [13], [14], [18] analyze the criminal behavior that occurred to compare it with previous stored crimes.

The other works like as [2], [4], [6], [16] provide techniques to determinate the places of criminal persons after studying the mechanisms of crime- excitation and geographical analysis of the data surrounding the crime.

The last works selected depend on finding a criminal person through finding his place, which is expected to be in [3], [10], [11], [12], [15].

All these works don't give any information about the early predication of the criminal and terroristic behavior, which we have studied. No other work in this area has been found.

Importance of the research and its goals

The treatment used in this research is to study the following situation: how do we get a prediction about the people who are going to commit acts of crime or terrorism, in order to find solutions to these crimes before committing them?

This strategy depends on determining some elements in our XML-document that play the major role in making the right decision. These elements are defined according to special experiments' results available all around the world, which we may change in accordance with circumstances to support our decision.

Research Methods and Materials:

Combined with the dynamic nature and complexity of criminal behavior, this sets the stage for successful data mining applications supported through XML- strategies.

With these XML-strategies the most important information formulate in form of XML- documents that depend on a number of attributes and elements classified under two main types. Those two types are the 'essential elements' and 'non-essential elements' for the processed documents.

Then we will define rules for our data mining strategies supporting the decision for early predication of criminal and terroristic behavior. These rules depend on simple type of decision support strategies. The Attributes and elements that we have defined in our XML- documents are fixed to help our data mining rules to give us the best information about the person that maybe criminal or terroristic.

The benefit from the application of this technology in schools and government institutions and security institutions in the communities will be illustrated by examples. Through these examples we can see how data mining plays an important role at operational, tactical and strategic levels of decision making. The main key to success has been a strong focus on simplicity.

Results and Discussion:

We begin now to explain and present the most important data according to our analyses after studying a behavior of some famous criminal or terroristic people since growing up to arrest. For these very important data we define a XML- schema with using of DTD. See [19], [20], [21], [22], [23], [24]. The DTD that we will have to define the XML- documents is presented as following:

1 -The DTD description for TEPTCP:

```
<!ELEMENT Persons (Person+)>
<!ELEMENT Person (PrName, PrClass+, PrEmployee,
                                    PrSex, PrAddress+)>
<!ATTLIST Person PrId CDATA #REQUIRED,
                                  PrAge CDATA #REQUIRED>
<!ELEMENT PrName (FirstName, MiddelName, LastName)>
<!ELEMENT FirstName (#PCDATA)>
<!ELEMENT MiddelName (#PCDATA)>
<!ELEMENT LastName (#PCDATA)>
<!ELEMENT PrClass (FamilyClass, FamilyHate,
                                    PrivateClass, PrChild)>
<!ELEMENT FamilyClass (#PCDATA)>
<!ELEMENT FamilyHate (#PCDATA)>
<!ELEMENT PrivateClass (#PCDATA)>
<!ELEMENT Child (#PCDATA)>
<!ELEMENT PrEmployee (#PCDATA)>
<!ELEMENT PrSex (male | female)>
<!ELEMENT male EMPTY>
<!ELEMENT female EMPTY>
<!ELEMENT PrAddress (City, Field, Street, Tel+, Mobile+)>
<!ELEMENT City (#PCDATA)>
<!ELEMENT Field (#PCDATA)>
<!ELEMENT Street (#PCDATA)>
<!ELEMENT Tel (#PCDATA)>
<!ELEMENT Mobile (#PCDATA)>
```

We have use the simplistic case of define a XML schema for a XML- document. The Attributes in this schema (PrId, PrAge) plays a big role to control XML-documents and for supporting our data mining strategy too like as the defined elements (the other one) in this XML- schema. Of course there is other kind to define this schema but we have it only to declare with which XML- documents we work.

Depending on this XML- schema we create the XML- documents for each person that we can have information about him. Then we define a decision tree depending on created XML- documents according on their defined XML- schema.

2- TEPTCP Level's strategy:

Before we give an example of suitable XML- document and its approach with using our Data mining rules for our Algorithm, we present the level strategy of our defined Attributes that we use to define the data mining strategy for our project.

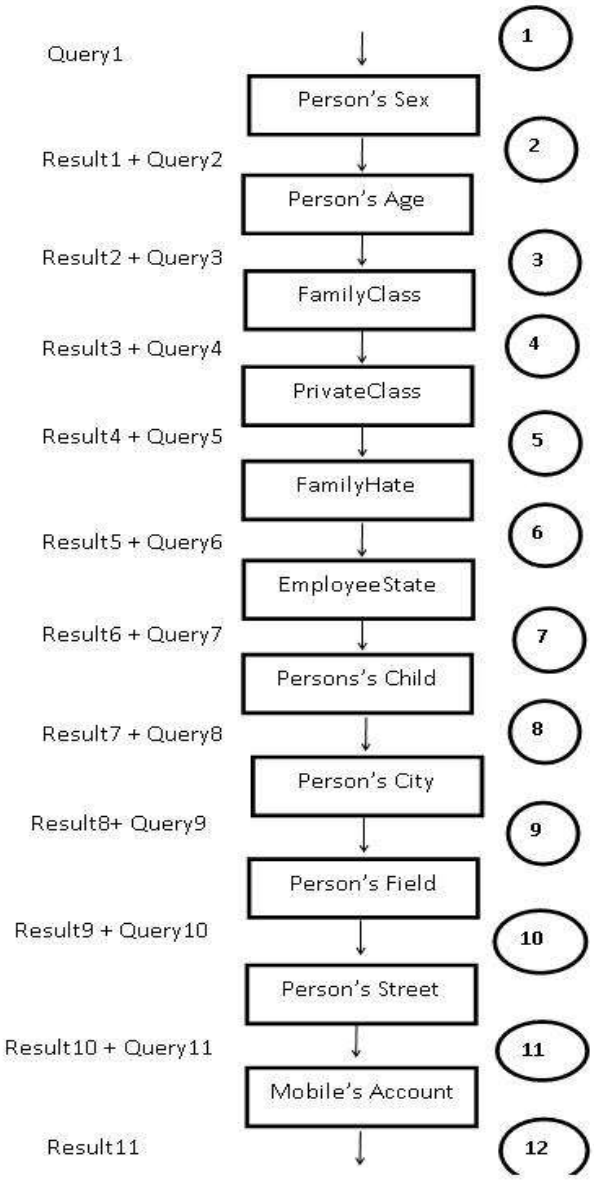

Figure (1): Level- strategy

For these 11 elements we use 11 queries. This means when we must have more elements to get better decision we need more queries, but this is no problem for our

technique because our process works in each level separately. As a consequence, we can use the parallel programming basis to give our project the best possibility to give the required result quickly.

3- TEBTCP Essential & Non- Essential Elements:

According on this Run-time level- strategy, we must define the parameters of each level, which must be unique and clear for our (Attribute) that appear in the illustrated level-strategy. We differentiate between essential and non- essential elements as follows:

Table 1: Contents of determinate elements

Type	Element	Content
Primary Essential	Sex	Either 'Male' or 'Female'
	Age	From 5 year until Death
	Family Class	Rich, Middle or Poor
	Private Class	Married or Not Married
	Family Hate	Very, Yes, No
	Person's Employee	HighEmp, MiddleEmp, LowEmp
	children	Yes or No
Non-Essential	City	Any Content
	Field	Any Content
	Street	Any Content
	Account of Mobile	Any Content
	Last Name	Any Content

The essential elements are fixed in our project according to our perspective and we can add more elements according to the usage. These elements play a very important role in our decision's tree. The non-essential elements contain some information that can play a decisional role for exact persons. The following table presents the dimension matrix and the operations used in this research.

Table 2: Initial dimension matrix and operations

Dimension / Operation	Sex	Age	Family C	Private C	Family h	Employee	Children	City	Field	Street	Mobile	LName
Very much Terrorist	×	×	×	×	×	×	×	×	×		×	
Very Terrorist	×	×	×	×	×	×	×	×	×		×	
Terrorist	×	×	×	×	×	×	×					×
unwell	×		×		×	×						×
Normal	×	×	×	×	×	×	×			×		

The symbol (×) explain that the suitable operation needs information from the suitable dimension to give the result.

4 -TEBTCP Decision Tree:

Now we present our decision's tree depending on the listed elements above. This tree can show all rules that we can define to find the best result in worst cases. We concentrate now only on the essential elements (parameters). The non- essential elements may help in other phases.

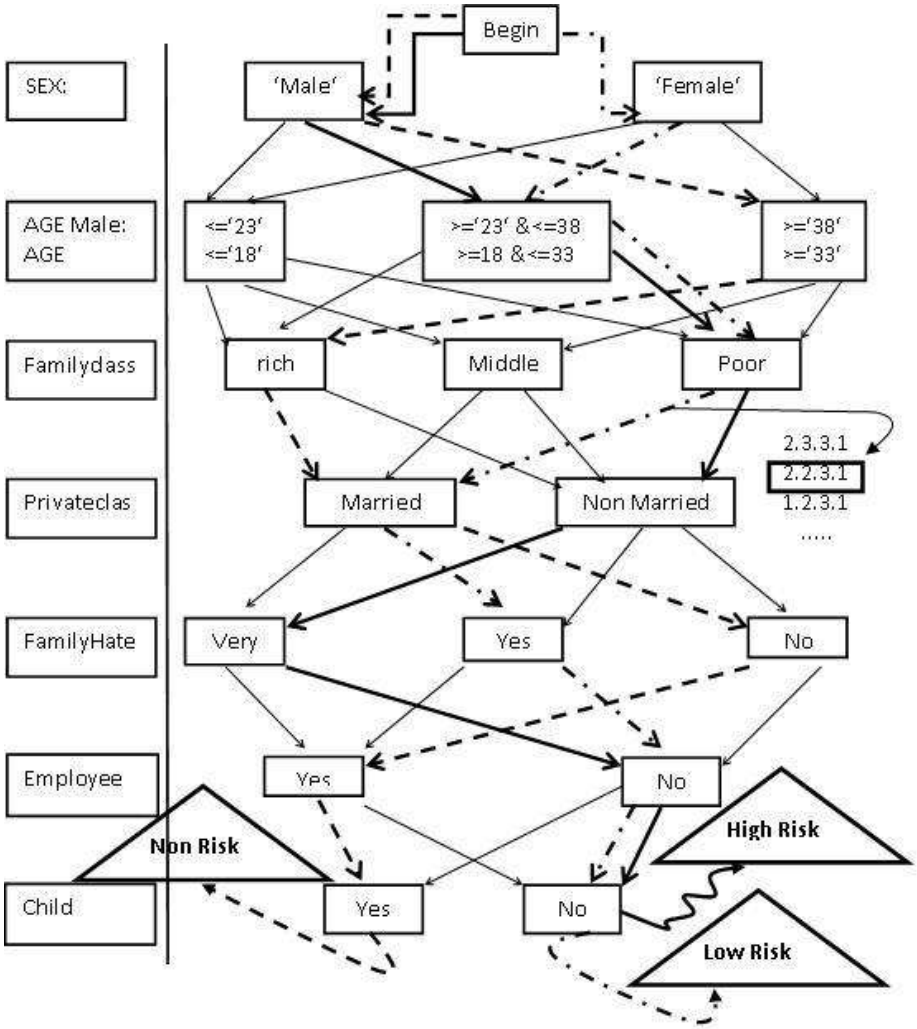

Figure (2): Decision Tree for essential elements

⟶ -,- -> ,- · - · -> **have the same meaning**

We have discrimination in this tree between the three types of lines connecting levels. Each line has a code to navigate into this tree. This means, through this code, we can trace the path to the specified value. Start coding the fields from left to right from the root of the tree, depending on the previous level's number.

For example, we find in figure (2) the path '2.2.3.1'. This means the value for the first attribute is the second one 'Female', then the value for the second attribute is the second one too '>=18 & <=33', for the third attribute the value is the third one 'poor' and for the fourth attribute the value is the first one 'married', etc.

In this tree, we find many directions that can be ordered for a person. This makes the decision more comprehensive and simpler to get it.

The process in our strategy can be worked sequential at the same level. This means we only need one and only one value to have in each level. This makes our process simple always. Through these possibilities, we can differentiate between 'High risk', 'Low risk' and 'Non- risk'. Then with support of Non- essential element in our level's strategy can be the decision more clearly (unique: in the best case).

This Tree can be still large and complicated but the execution of the strategy for the created tree stays uncomplicated. After the creation of the desired tree begin we to define the data mining rules depending on the created tree.

5- TEPTCP Data Mining Rules:

In this article, we will define some rules depending on essential elements that can help the mechanism to predict the road, which could be the route of the person, and then we define other rules depending on the non- essential elements. In some cases, this strategy can be very useful to predict the presumable person exactly, no matter how big and complicated the tree will be.

1- First Rule:

If 'sex= male' AND 'age <=23 OR [age>=23 AND age<=38'
AND 'FamilyClass = Poor' AND 'PrivateClass = Non Married'
AND 'FamilyHate = Very' AND 'Employee = NO'
AND 'Child = NO'
Then Risk = 'High'

According to this rule, we note that if the gender of person is male, his age is greater than 23 and less than 38, his family-class is poor, is not married and has big hate for his family, is not employee and has no children, then this person is on the way to be a terrorist.

2- The second Rule:

If 'sex= male' AND 'age >=38'
AND 'FamilyClass = Rich' AND 'PrivateClass = Married'
AND 'FamilyHate = No' AND 'Employee = Yes'
AND 'Child = Yes'
Then Risk = 'Non'

According to this rule, we note that if the gender of person is male and his age is greater than 38, his family-class is rich and is married, has no hate for his family, is employee and has children. then this person is not on the way to be a terrorist.

3- The third Rule:

If 'sex= Female' AND '[age >=18 AND age<=33]'
AND 'FamilyClass = Poor' AND 'PrivateClass = Married'
AND 'FamilyHate = yes' AND 'Employee = No'
AND 'Child = No'
Then Risk = 'Low'

According to this rule, we note that if the gender of person is female and his age is greater than 18 and less than 33, his family-class is poor and is married, has hate for his family and is not employee and has not children, then this person is on the way to be terrorist.

These are the three rules for three different cases without using the Non- essential elements. We can surely define more efficient rules according to each application depending on its society. In some cases, we can find some persons that must belong to the 'High risk'-, 'Low risk' or 'Non risk' class, but they are not so. In this case, we add other elements for our strategy that can be helpful depending on suggestions of the expert scientists.

Note: We did not use the non- essential elements, because we do not want to demonstrate information about cities or persons' last names to avoid misunderstanding for these cities and families. In this part of our project, we need sometimes full-text research techniques. See [25], [26],[27].

How we can support our project with XML and XQuery technologies, can be found in the next section.

6- Application of the previous strategy (TEBTCP) practically

Example:

To be the following XML document:

```
<Persons>
<person PrId=’321’ Prage=’39’>
        <PrName>
                <FirstName>DDD</firstName>
                <MiddleName>EEE</MiddleName>
                <LastName>FFF</LastName>
        </PrName>
        <PrClass>
                <FamilyClass>Rich</FamilyClass>
          <PrivateClass>Married</PrivateClass>
                <FamilyHate>No</FamilyHate>
                <PrChild>Yes</PrChild>
        </PrClass>
        <PrEmployee>Yes</PrEmployee>
        <PrSex>Male</PrSex>
        <PrAddress>…</PrAddress>
  </person>
  <person PrId=’123’ Prage=’27’>
        <PrName>
                <FirstName>AAA</firstName>
                <MiddleName>BBB</MiddleName>
                <LastName>CCC</LastName>
        </PrName>
        <PrClass>
                <FamilyClass>Poor</FamilyClass>
        <PrivateClass>Non Married</PrivateClass>
                <FamilyHate>Very</FamilyHate>
                <PrChild>No</PrChild>
        </PrClass>
        <PrEmployee>No</PrEmployee>
        <PrSex>Male</PrSex>
        <PrAddress>…</PrAddress>
  </person>
  <person PrId=’231’ Prage=’30’>
        <PrName>
                <FirstName>GGG</firstName>
                <MiddleName>HHH</MiddleName>
                <LastName>III</LastName>
        </PrName>
        <PrClass>
                <FamilyClass>Poor</FamilyClass>
          <PrivateClass>Married</PrivateClass>
                <FamilyHate>Yes</FamilyHate>
                <PrChild>No</PrChild>
        </PrClass>
```

```
            <PrEmployee>No</PrEmployee>
            <PrSex>Female</PrSex>
            <PrAddress>…</PrAddress>
        </person>
    </Persons>
```

All 'PrAddress'- elements in our XML- document must be entered with specified data.

This XML- document must be stored in our database. We give it the name 'persons.xml'. The suitable XQuery for our XML document that can give us the concerned person if the risk is 'high', 'Low' and 'Non' can be written as following: (all under show queries can be written otherwise, according to priorities of focused elements.)

1- XQuery for the first rule shows as following:

```
let $p := doc('person.xml')\\Persons
for $pp in $p\Person
where $pp\PrSex ='Male' and $pp\@PrAge <=23 or ($pp\@PrAge>=23 and $pp\@PrAge<=38)) and $pp\PrClass\FamilyClass = 'Poor' and $pp\PrClass\PrivateClass ='Non Married' and $pp\PrClass\FamilyHate = 'Very' and $pp\PrClass\PrEmployee = 'NO' and $pp\PrClass\PrChild = 'NO'
  return <HighRisk>$pp\PrName</HighRisk>
```

2- XQuery for the second rule shows as following:

```
let $p := doc('person.xml')\\Persons
for $pp in $p\Person
where $pp\PrSex ='Male' and $pp\@PrAge <=38 and $pp\PrClass\FamilyClass = 'Rich' and $pp\PrClass\PrivateClass = 'Married' and $pp\PrClass\FamilyHate = 'No' and $pp\PrClass\PrEmployee = 'Yes' and $pp\PrClass\PrChild = 'Yes'
return <NonRisk>$pp\PrName</NonRisk>
```

3- XQuery for the third rule shows as following:

```
let $p := doc('person.xml')\\Persons
for $pp in $p\Person
where $pp\PrSex ='Female' and $pp\@PrAge >=18 and $pp\@PrAge<=33 and $pp\PrClass\FamilyClass = 'Poor' and $pp\PrClass\PrivateClass = 'Married' and $pp\PrClass\FamilyHate = 'Yes' and $pp\PrClass\PrEmployee = 'No' and $pp\PrClass\PrChild = 'No'
return <LowRisk>$pp\PrName</LowRisk>
```

All these queries can be written in other form depending on the priorities of the elements highlighted in the decision-making. The results of these queries must be stored in extra XML- document and then warnings must be given for security workers. For example the result of the first query is:

```
<HigRisk>
<PrName>
            <FirstName>GGG</firstName>
            <MiddleName>HHH</MiddleName>
            <LastName>III</LastName>
        </PrName>
</HigRisk>
```

We save this document with the name 'highrisk.xml'.

Conclusions and recommendations:

This project is very important for parents, society and country security. This project can give the security reports more discipline. These reports can take more usefulness for security, police and people. This technique can be completed with additional elements and simple algorithms. It needs much useful information and only one and many discipline for workers on this project.

The use of XML and XQuery make the transfer of data between many resources simpler and clear. We can use our technique in distributed systems and with attractive design depending on practice views and mythologies. With the non- essential defined element, we need for the full-text research the information retrieval technique too. In some cases, we can avoid these elements. These can give more exact decisions. We can study more parametric and non-parametric Data mining algorithms to achieve the best solutions.

References:

[1]VAN DER VEER R.; ROOS MSC H. T.; VAN DER ZANDEN A. MSC, *Data Mining for intelligence led policing*, Proceedings of the 15TH ACM SIGKDD International conference on Knowledge Discovery And Data Mining 2005.

[2]RATCLIFFE J. H.. *Intelligence- led Policing*, Australian Institute of Criminology, Canberra, Australia, 2003, 248.

[3]ADDERLY R.; MUSGROVE P.B., *Data Mining Case Study: Modeling the Behavior of Offenders Who Commit Serious Sexual Assaults*, Proceedings of the seventh ACM SIGKDD International conference on Knowledge Discovery And Data Mining, 2001. 215-220.

[4]McCue C. *Data Mining and predictive analysis: Intelligence gathering and crime analysis*, Butterworth-Heinemann. 2007

[5]MENA J. *Investigative Data Mining for security and criminal detection*, Elsevier Science (USA), 2003. 272.

[6]BROWN D. E. *The Regional Crime Analysis Program (RECAP): A framework for mining data to catch criminals*, University of Virginia. 1998. 2828- 2853

[7]BRUIN J. S.; COCX D.; KOSTERS T.K.; LAROS J. F.J. KOK J. N., *Data Mining Approaches to Criminal Career Analysis*, Proceedings of ICDM 06. 2006. 171- 177.

[8]UYL M. J. *Representing Magnitude by memory Resonance*, Proceedings of the 6th Annual Conference of the Cognitive Science Society, 1986, 63- 71.

[9]BORG I.; GROENEN, *Modern Multidimensional Scaling: Theory and applications* (2nd ed.), SpringerVerlag NewYork.2005. 614.

[10] FARRELL G. *Progress and Prospects in the prevention of repeat victimization*. 2005. 816.

[11] ECK. J. E.; CHAINEY S.; CAMERON J. G.; LETINER; R.E. *Mapping crime: Understanding hot spots,* M. & Wilson,Technical report 209393, National Institute of Justice. 2005. 79.

[12] CHAINEY S.; RATCLIFFE J. *GIS and Crime Mapping, Mastering GIS: Technology, applications and management*. West Sussex, England: Wiley. 2005. 407.

[13] CLARKE R. V.; M. Felson, *Eds. Routine Activity and Rational Choice (Advances in Criminological Theory).* Transaction Publishers, New Brunswick (USA). 2004. 432.

[14] BOWERS K. J.; JOHNSON S. D.; PEASE K. *Prospective hot-spotting: The future of crime mapping?*, Br. J. Criminal 44(5), 2004. 641-658.

[15] KOHONEN T. *Self-organisation and associative memory*, Springer series in information science, Vol8. Springer Verlag, New York, USA.1984. 312.

[16] ROSSMO D. K. *Geographical profiling*, Boca Raton, Fla, CRC Press. 2000.347.
[17] FREEMAN L. *The Development of social network analysis*, Vancourver, CA: Empirical Press. 2004. 218
[18] KOHONEN T. *Self-Organizing maps*, Springer series in information Sciences, Vol30. Springer, Berlin, Heidelberg, New York. 2001. 236.
[19] BEERI C.; MILO T., *Schemas for integration and translation of structured and semi-structured data*. In Catriel Beeri and Peter Buneman, editors, Database Theory - ICDT '99, 7th International Conference, Jerusalem, Israel, January 10-12, 1999, Proceedings, volume 1540 of Lecture Notes in Computer Science, Springer, 1999. 296–313.
[20] BOAG S.; CHAMBERLIN D.; FERNANDEZ M. F.; FLORESCU D.; ROBIE J.; SIMEON J., *XQuery 1.0: An XML Query Language* . Technical report, W3C Working Draft 15 November, 2002. 207.
[21] BIRON P. V.; MALHOTRA A.; *XML Schema Part 2: DataTypes*. W3C Candidate Recommendation, May 2001
[22] DOBBIE G.; JACKY W.; WAN W., *Extracting Association Rules from XML Documents using XQuery*. WIDM03, November7-8, New Orleans, Louisiana, USA, ACM 1-58113-725-7/03/0011, 2003. 94–97
[23] MARIAN A.; SIMEON J., *Projecting XML Documents*. Proceedings of the 29th VLDB Confernence, Berlin, Germany, 2003. 1-12
[24] FALLSIDE D. C.; *XML Schema Part 0: Primer*. W3C Candidate Recommendation, October 2000. 24:1–55.
[25] BAZEA -YATES R., RIBERIRO-NETO B., *Modern Information Retrieval*. New York. 1999. 513.
[26] CARSTEBSEN K., et al. (Hrsg.) *Computerlinguistik und Sprachtechnologie*. Muenchen. 2004. 642.
[27] FERGUSON; LUCENE J.: *a tutorial introduction to ful-text indexing in Java*. 2005, 10 Junín 2010.
http://www.jroller.com/page/wakaleo/?anchor=lucene_a_tutorial_intorduction_to 2005.