

The prediction of terrorist threat on the basis of semantic association acquisition and complex network evolution

Andrzej Najgebauer, Ryszard Antkiewicz, Mariusz Chmielewski, and Rafał Kasprzyk

Abstract—In this paper, which is the continuation of an *MCC 2006 Conf.* publication by the same group of authors, we propose a concept of early detection of terrorist action preparation activities. Our ideas rely on semantic and complex networks to extract useful information for terrorist threat indication. Presented methods will be used as a core framework for Early Warning System.

Keywords—*Early Warning System, semantic network, ontology, complex network.*

1. Introduction

Detecting terrorist threats requires a large spectrum of data which in many cases are collected from various sources. The process of unification, fusion and interpretation of the collected data is crucial due to data redundancy and specially to enable accurate predictions. For knowledge representation we propose a semantic network based on an ontology data model. Using semantic graph as a storage for facts and events we have been able to develop a method for indirect association acquisition which allows us to pinpoint new relationships in our knowledge base to indicate possible threats. The algorithms are designed as two separate groups which are aimed at rule/ontology based inference and graph structure analysis. Both groups provide different approaches, which allow more accurate possible threat extraction. Implementation of the concept is based on the known standard for semantic data representation web ontology language (OWL) and resource description framework (RDF), while for inference engines we used the Java based frameworks JENA and JADE.

In this paper we will introduce the design goals of ontology used in our work, methods of filtering unreliable data and most of all the concept of dynamic graph analysis using an agent based environment. Considering the large scale of the data set and the complexity structure analysis algorithms, we were forced to provide additional modifications which in essence are filtering ontologies. Using this method we reduce the size of semantic network dividing the original semantic graph in two parts and only one described in a filtering ontology is the input for algorithms.

We focus special attention on research in complex networks. These kinds of networks have scale free, small world and clustering features what make them accurate models of spontaneously growing networks such as social networks and in particular terrorist organizations. We are able to

transform any semantic network into a set of complex networks by choosing the ontology which is important for us at the moment of analysis.

The main problem is one of choosing the way to represent the structure of interest in a complex network and what set of measurements of the topological features are the best network characterization described by the net components (nodes and edges).

Of particular interest is the relationship between the structure and dynamics of complex networks. We are convinced of the importance of measuring the structural properties of evolving networks in order to characterize how the connectivity of the investigating structures being investigated changes in time. Network measurements are therefore essential in our investigation. We consider using the following significant characteristics of a given network: clustering coefficient, average path length, shortest longest path and preferential attachment. Now we are able to calculate these characteristics of complex network and observe its changing in time. We are convinced that the vector of significant network characteristics has to be much more numerous than the four we have proposed. The question is, how can the non-stability of the characteristics mentioned above (collected measures) be viewed as a factor that indicates an abnormal state of the system (e.g., increasing terrorism activity) modeled by complex networks?

In a previous paper [1] we introduced the Early Warning System (EWS) concept. The EWS is a simulation-based diagnostic support tool, with its associated algorithms, that realises the following processes:

- Collecting information relevant to terrorism threat estimation and intelligence data analysis from:
 - primary threat factors determination,
 - aggregated threat factors (causative and executive) determination,
 - threat coefficient estimation,
 - possible goals of terrorist attack identification.
- The analysis and simulation, using the collected information in order to: predict the terrorism threat over long periods of time, predict the stability of the threat factors and the detect when pre-determined threat factor thresholds have been exceeded.
- The visualization of EWS output for potential users.

2. Association acquisition

As mentioned in [1] methods used for terrorist threat recognition and evaluation are the following [10, 12, 14]:

1. Graph path finding algorithms.
2. Rule based inference engine – inference algorithms working with semantic information could provide an optional source of indirect association (building new knowledge).
3. Graph similarity algorithms – identifying crucial patterns in semantic networks which are significant for the asymmetric conflict scenarios.

The process of data acquisition for the building of the knowledge base is executed by program agents which re-view external data sources.

Our proposed method of building associations contains several stages [2]:

- Designing a generic dynamically changing ontology allowing flexible information representation.
- Designing the mechanisms for knowledge acquisition for building of the semantic network (also designing intelligent agents who provide the database search algorithms).
- Defining parameters which will allow analysis of the knowledge in the network (dependency analysis, clustering connected to nodes concerning the base ontology).
- Definition of particular ontologies, which are used as filters for elimination of unnecessary links in the net.

To define specific paths let us introduce the following description [2]:

- Paths – heuristic route search algorithms, are based on problem size reduction using the reference to the ontology analysis but not the semantic network itself. Metamodel usage allows decreasing quantity of nodes and links to be analyzed by the algorithm. The linking of nodes to the calculated route (building association) is achieved using the depth-first algorithm, considering the currently analyzed node and the ontology template which gives the information of all valid links to other nodes.
- Intersecting paths – using the definition of paths on the semantic graph the algorithm is searching for two paths, containing intersecting links which connect nodes in those paths.
- Isomorphic paths – are based on the algorithm of finding two paths which are isomorphic, that means that we have to find two pairs of nodes with such paths in the network that any of the nodes from one path is a part of the other path.

The idea of our system's activity, depicted in Fig. 1 was originally given in [19].

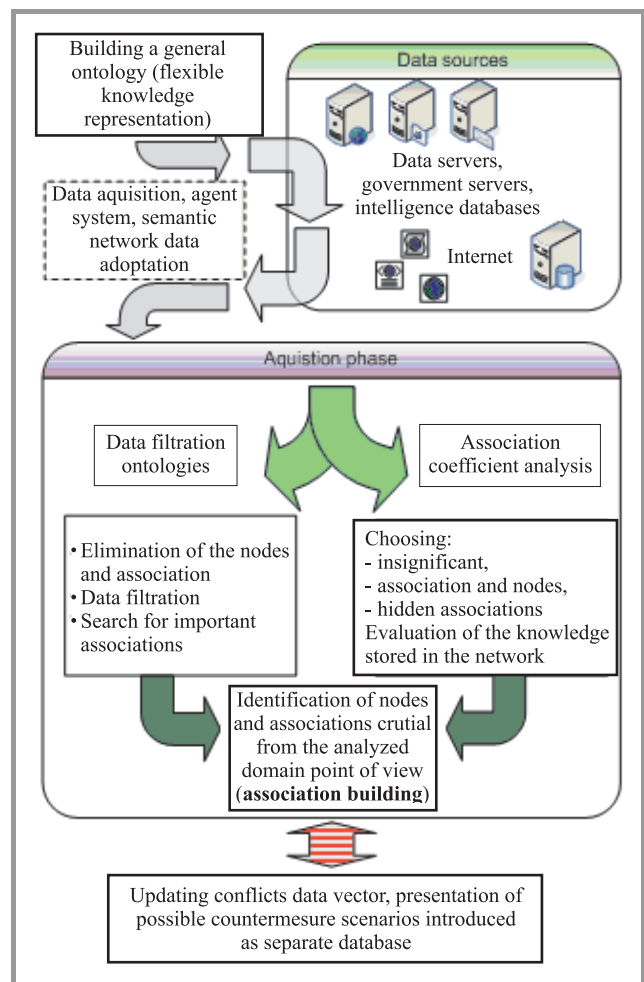


Fig. 1. Idea of terrorist threat evaluation based on semantic model of threat [19].

3. Acquiring new knowledge

Acquiring new knowledge in semantic network is based on introducing new nodes and links between nodes [2]. This can be achieved in two ways: using the analysis of the structure of the semantic network, and inference engines. Inference algorithms can be implemented as:

- General logic based inference engine – where there are two main aims, first order logic, higher order logic and description logic. First order logic (FOL) are mechanisms which are very efficient but computationally not tractable for large amounts of data and axioms. Higher order logic based engines, however, are able to track the inference route but they require a lot of resources to achieve the task.
- Solving algorithms – are specialized algorithms, often small size, designed to provide a solution in one distinct problem. Problem solving methods (PSM) define which actions in the whole inference process need to be executed and how the control flow in such algorithm should look like (considering the control of the subtasks).

The idea was implemented using standards for semantic data representation RDF and OWL. For the inference engine and tools allowing representation of semantic models we use the JENA OpenSource API (see the description of EWS environment).

4. Complex network evolution

Complex networks (CN) with scale free, small world and clustering features are accurate model of spontaneously growing networks such as: Internet, WWW, social networks [3, 5, 6, 8]. Our work has demonstrated that in some cases we can use such complex network to automatically detect and or estimate some aspects of a terrorist organization by treating these as a special kind/type of social network.

This part of our work is strongly connected with social networks. Social network analysis is a collection of mainly statistical methods to support the study of communication relations in groups, kinship relations, or the structure of behavior, to mention a few application areas. This methodology assumes that the way the members of a group can communicate with each other affects some important properties of that group. We have applied social network analysis in anti-terrorism applications and indicate both its usefulness and some of its limitations when using it as a quantitative method for situation awareness and decision-making in law-enforcement applications. Understanding nested connections across a known set of individuals or organizations is one example of social network analysis. Since not all people who have had contacts with a terrorist are criminal themselves, there is a need for techniques which can filter out those who have frequent contacts with known or suspected individuals, or with any member of a known or suspected group of terrorists from a large database of contacts. Such people become more or less suspect themselves, thereby potentially spreading the suspicion to even more individuals.

One of the important issues is connected with the question how can we automatically estimate which people among a very large community, who have been “transitively” in contact with each other, need to be investigated further and who do not. We explore such methods of social network analysis using a complex network as a model of a terrorist organization.

We are able to transform any semantic network into a set of complex networks by choosing the ontology which is important at the moment of analysis.

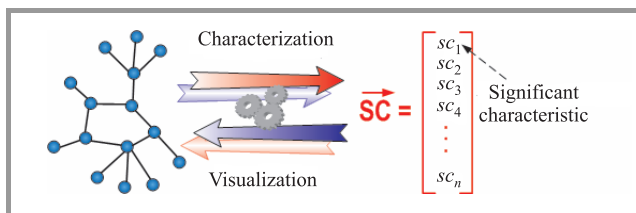


Fig. 2. Mapping from CN into a feature vector.

The main problem is to choose how to represent the structure of interest as a complex network and what set of measurements of the topological features are the best network characterization described on the net components (nodes and edges).

We are convinced that the vector of significant characteristics (SC) have to be numerous (Fig. 2). We consider using following significant characteristics of a given network:

- $sc_1 \equiv C$ – clustering coefficient,
- $sc_2 \equiv L$ – average path length,
- $sc_3 \equiv l$ – shortest longest path,
- $sc_4 \equiv D$ – diameter,
- $sc_5 \equiv \langle k \rangle$ – average node degree,
- $sc_6 \equiv k_{max}$ – maximum node degree,
- $sc_7 \equiv P(k)$ – node degree distribution,
- $sc_8 \equiv PA$ – preferential attachment.

Another problem we consider is how to use the obtained structural properties (measurements) in order to identify different categories of structures, which is directly related to the area of pattern recognition. Each class of networks

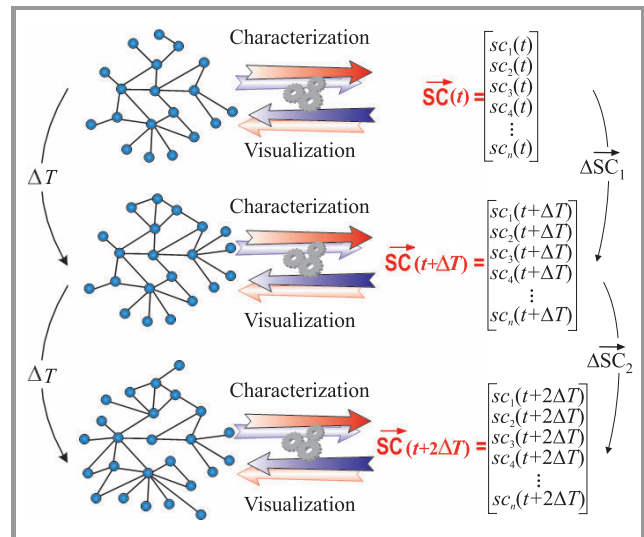


Fig. 3. Evolution of a given network and result changes of a feature vector.

presents specific topological features which characterize its connectivity and highly influence the dynamics of processes executed on the network. The analysis, discrimination, and synthesis of networks (in particular complex networks) therefore rely on the use of measurements capable of expressing the most relevant topological features.

We focused attention particularly on the relationship between the structure and dynamics of complex networks. We are convinced of the importance of measuring the structural properties of evolving networks in order to characterize how the connectivity of the investigated structures changes in time. This can help to identify some odd events in modeled system.

The vector of significant characteristics should be updated at each ΔT along the network growth/decline. Figure 3 shows instances of the evolving network and respective measures. This implies the very important question of how to choose the most appropriate measures for a given system. The answer must reflect the specific interest and it is still an open question under our investigation.

Network measurements are therefore essential in our investigation. We can calculate significant characteristics of a complex network as mentioned above and observe its change in time. We intend to test how the non-stability of characteristics mentioned above (collected measures) can be viewed as a factor that show an abnormal state of the system (e.g., increase in terrorism activity) modeled by complex network.

Important related issues covered in our work comprise the representation of the evolution of complex networks in terms of trajectories in several measurement spaces, the analysis of the correlations between some of the most traditional measurements, perturbation analysis, as well as the use of multivariate statistics for feature selection and network classification.

Figure 4 presents the sample trajectory defined in one of the possible measures (phase) spaces involving three possible significant characteristic $sc_i(t)$, $sc_{i+1}(t)$ and $sc_{i+2}(t)$. In such a way, the evolution of the network can be investigated in terms of a trajectory in a “phase space” using chaos theory for example.

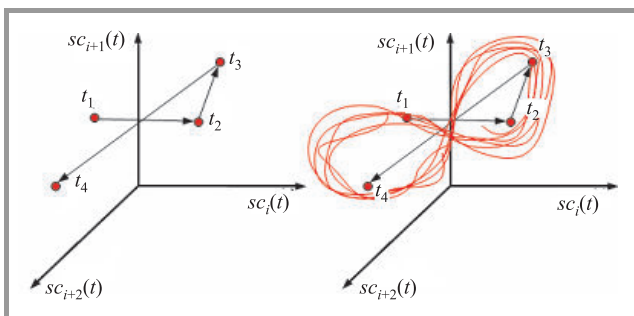


Fig. 4. Trajectories in the “phase space” defined on the basis of feature vector changing.

The ultimate purpose of the project is to simulate and additionally visualize various scenarios of attack and defence to investigate bottlenecks in security system. Visual representation of information can be used to demystify data and reveal otherwise hidden patterns by leveraging human visual capabilities to make sense of completely abstract information (see the EWS environment description).

5. The EWS environment

Sections of EWS portal prototype consist of the following parts.

Logging users – each user needs to register himself and log in each time he wants to make use of portal functionality. Registering user consist of two phases. First, the user

submits his personal data. In the second phase, performed by the administrator using the portal, the user permissions are submitted to each service (semantic network analysis, complex network analysis, etc.).

Registering really simple syndication (RSS) – the portal obtains multiple news reports from different news sources to show a ticker console (text scroller) containing news reports on any registered terrorist activity or terrorist report. The RSS database is flexible and is updated in real-time. The RSS sources stored on server can be extended as needed (Fig. 5).

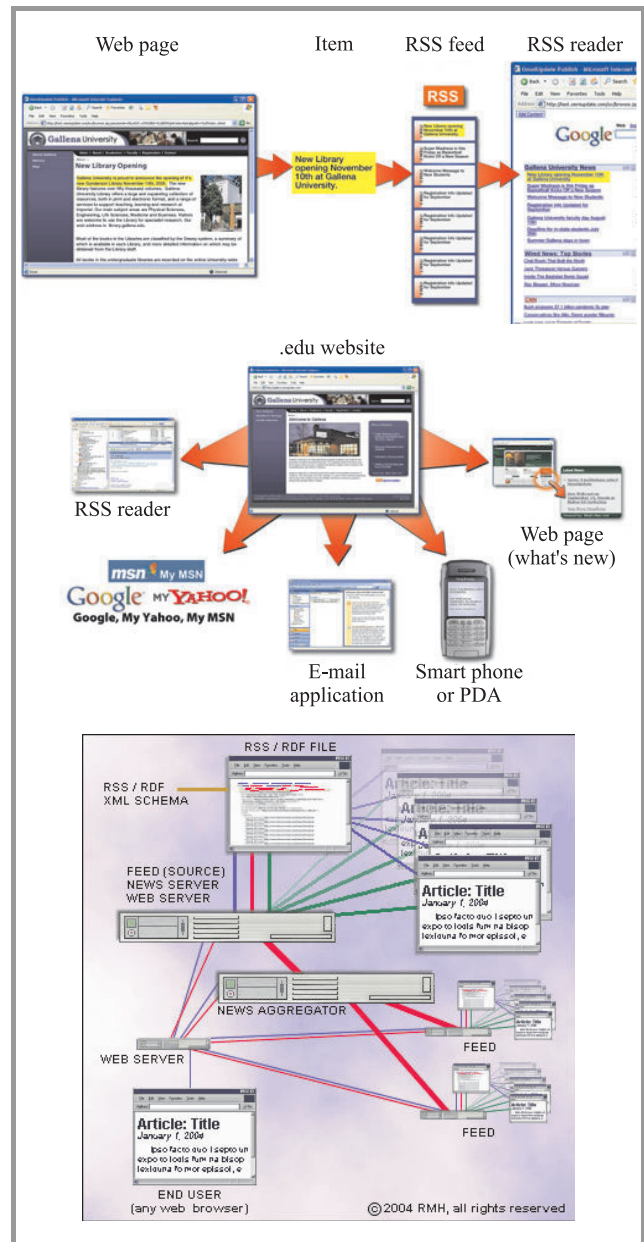


Fig. 5. Concept of aggregation RSS data between data sources.

Complex networks – link analysis – are tools for visualizing associations between stored data (e.g., terrorist social activity – relation “knows”, “contacts”). The data shown in figures have been taken from the September 11th attack.

We propose an additional example as we have extracted the log from our mail server and run several graph algorithms such as maximum clique or k -clique algorithm to extract groups of users communicating with each other.

Semantic network analysis – this category has been divided in two parts. First one visualizes the ontology we use to represent all stored data and relations in our semantic graph (Fig. 6). This part directly presents the part of the whole terrorist ontology as the full model is large and it would be difficult to present. The other part is our proposal for a method for indirect data association searching. For this part we use the emerging standards for semantic data representation RDF and OWL. For the inference

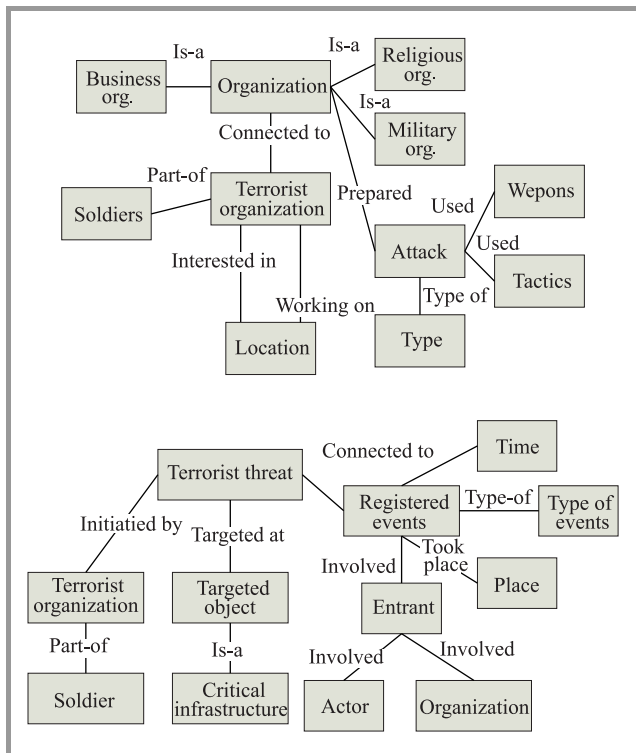


Fig. 6. Visualizing the ontology.

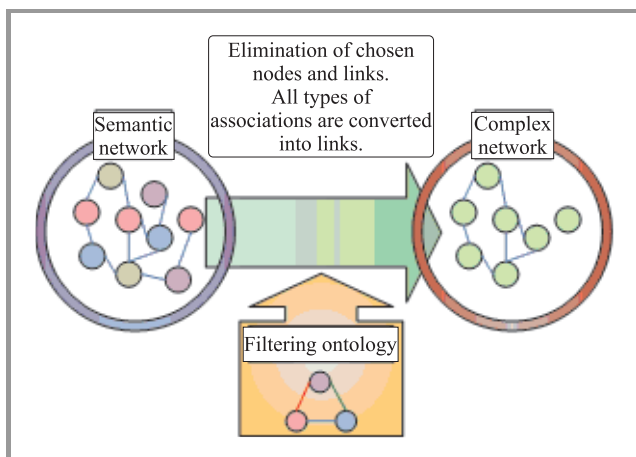


Fig. 7. The transition between semantic network and complex network using ontology filtering.

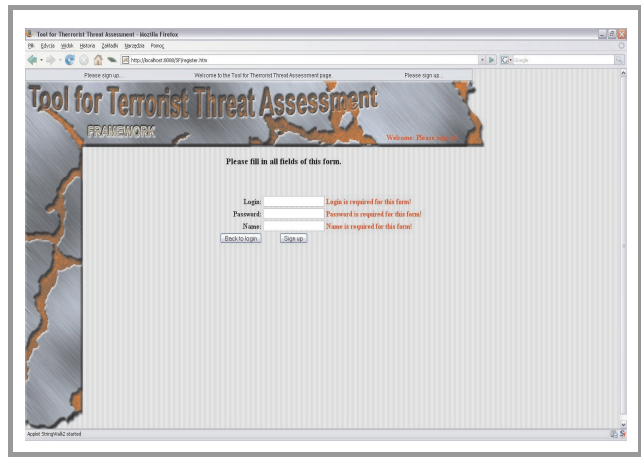


Fig. 8. User registration for EWS portal.

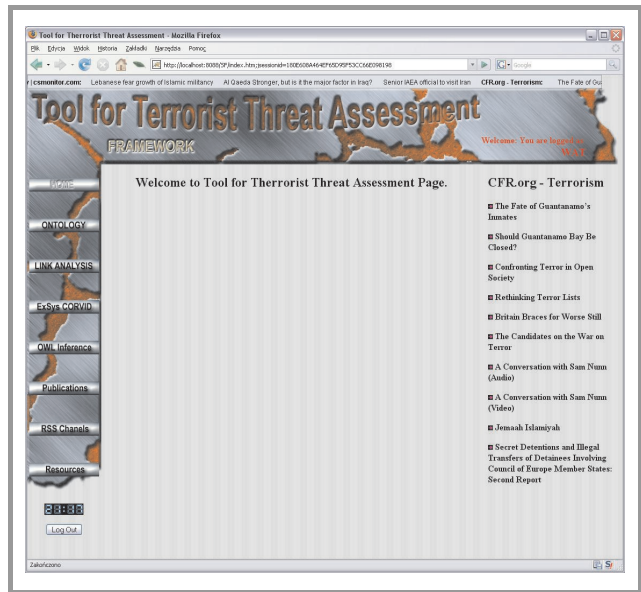


Fig. 9. Homepage of EWS portal.

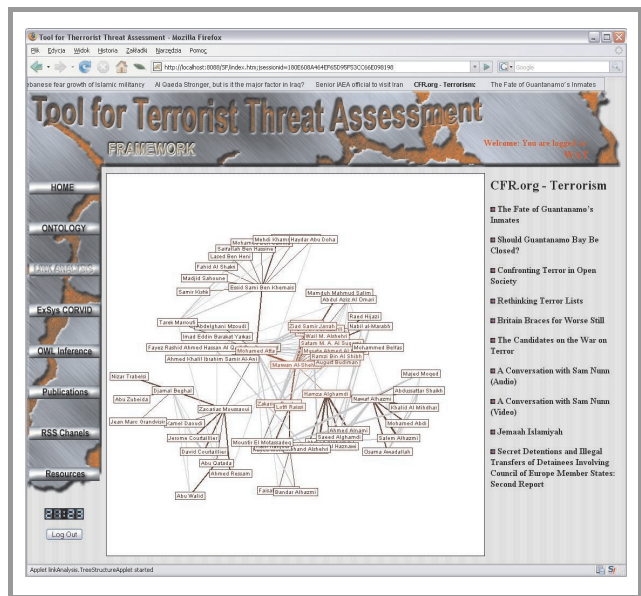


Fig. 10. Example of link analysis for complex networks – data from 11th September attack.

engine and tools allowing representation of semantic models we use JENA OpenSource API.

The semantic network filled with the collected data, for the further analysis is transformed into the complex network. The transformation is conducted as a filtering process based on a set of filtering ontologies. We use them as filters to reject unneeded data and extract only those relations that are useful for link analysis. The scheme of the process is shown in Fig. 7.

Additional functionalities provided include support for rule based inference with ExSys Corvid which allows providing a questionnaire for the analyst which will guide him through the process of collecting all information on registered data/actions. The algorithm provides, based on

the input data, a result which represents the probability levels for stored scenario attacks.

For the EWS system we have also provided a regression model for threat assessment. This was described in presented publications [1], and is available in the portal. It requires a path to the defining database content and set of primary and secondary factors. Using neural networks and clustering techniques it is able to evaluate a set of threat coefficients for the current set of stored crises. The windows for different pages of the EWS portal are illustrated in Figs. 8–12.

6. Conclusion

In this paper we presented methods and a prototype of a EWS for terrorist threat identification which can be developed into a professional analysis tool. The solutions presented in the paper are applied in the project of Crisis Management System for big agglomeration.

Acknowledgements

This work was partially supported by grant “Research Project no. PBZ-MIN/011/013/2004” and NATO RTO MSG 026.

References

- [1] A. Najgebauer, “A concept of simulation based diagnostic support tool for terrorism threat awareness”, in *Model. Simul. Adm. NATO's New Exist. Milit. Req. Conf.*, Koblenz, Germany, 2004.
- [2] D. J. Watts and S. H. Strogatz, “Collective dynamics of “small-world” networks”, *Nature*, vol. 393, pp. 440–442, 1998.
- [3] B. A. László and A. Réka, “Emergency of scaling in random networks”, *Science*, vol. 286, pp. 509–512, 1999.
- [4] S. H. Strogatz, “Exploring complex networks”, *Nature*, vol. 410, pp. 268–276, 2001.
- [5] B. A. László and A. Réka, “Statistical mechanics of complex networks”, *Rev. Mod. Phys.*, vol. 74, pp. 47–97, 2002.
- [6] M. E. J. Newman, “Models of the small world: a review”, *J. Stat. Phys.*, vol. 101, pp. 819–841, 2000.
- [7] M. E. J. Newman, “The structure and function of complex networks”, *SIMA Rev.*, vol. 45, no. 2, pp. 167–256, 2003.
- [8] W. Xiaofan and C. Guanrong, “Complex networks: small-world, scale-free and beyond”, *IEEE Circ. Syst. Mag.*, vol. 3, no. 1, pp. 6–20, 2003.
- [9] V. Krebs, “Mapping networks of terrorist cells”, *Connections*, vol. 24, no. 3, pp. 43–52, 2002.
- [10] J. Golbeck, A. Mannes, and J. Hendler, “Semantic Web Technologies for Terrorist Network Analysis”. IEEE Press, 2006.
- [11] M. Barthelemy, E. Chow, and T. Eliassi-Rad, “Knowledge representation issues in semantic graphs for relationship detection”, UCRL-CONF-209845, <http://www.edmondchow.com/pubs/>
- [12] A. Mannes and J. Golbeck, “Building a terrorism ontology”, University of Maryland, College Park, 2005.
- [13] M. Steyvers and J. B. Tenenbaum, “The large-scale structure of semantic networks: statistical analyses and a model of semantic growth”, *Cogn. Sci.*, vol. 29, pp. 41–78, 2005.
- [14] E. Suzikov and D. Soshnikov, “Using dynamic ontologies based on production-frame knowledge representation for intelligent web retrieval”, in *Worksh. Comput. Sci. Inform. Technol.*, Patras, Greece, 2002.
- [15] J. F. Sowa, “Semantic networks”, <http://www.jfsowa.com/pubs/semnet.htm>

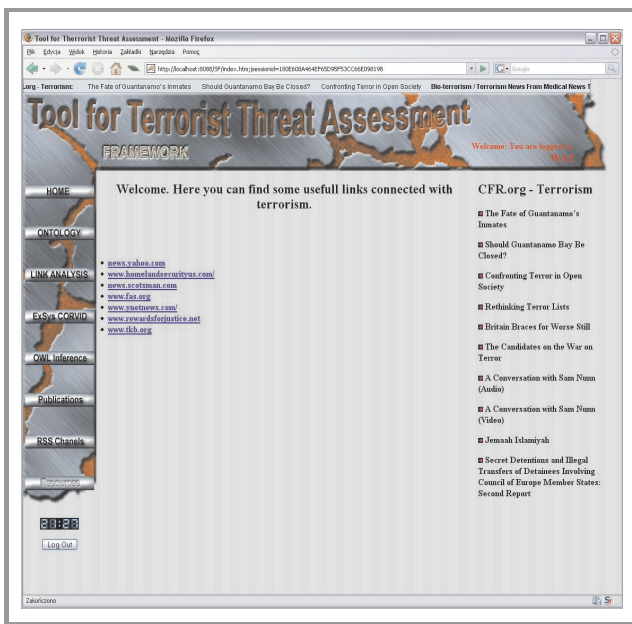


Fig. 11. Database of hyperlinks to other resources connected with the domain of counterterrorism.



Fig. 12. Quick RSS content viewer.

[16] JENA web semantic framework, <http://jena.sourceforge.net/>
 [17] Resource definition framework homepage, <http://www.w3.org/RDF/>
 [18] "Generic Early Warning Handbook", NATO/EAPC/PFP, 2001.
 [19] A. Najgebauer, R. Antkiewicz, M. Chmielewski, and R. Kasprzyk, "Terrorist threat identification using semantic associations and complex networks", in *Proc. MCC 2006 Conf.*, Gdynia, Poland, 2006.



Andrzej Najgebauer is the Dean of Cybernetics Faculty at Military University of Technology (MUT), Warsaw, Poland. He has M.Sc. degree in computer science (MUT, 1981), Ph.D. in computer sciences, (MUT, 1988), Certificate, Doctor of Science in decision support systems (Warsaw University of Technology, 1999). His

work is connected with modeling and simulation, designing of military DSS, conflict analysis, war games, exercise and training systems. Leadership of new Polish Army Simulation System for CAXes. He is a member of IFORS and Polish Society of ORSA, Polish Society of Computer Simulation. Polish representative of RTO/NMSG and activity leader in subject of Early Warning Systems for terrorist crisis. Leader of projects on security research.

e-mail: anajgebauer@wat.edu.pl
 Faculty of Cybernetics
 Military University of Technology
 Gen. S. Kaliskiego st 2
 00-908 Warsaw, Poland



Ryszard Antkiewicz received his M.Sc. in 1989 and Ph.D. in 2004 from Military University of Technology, Poland. He has worked in Cybernetics Faculty of Military University of Technology since 1984. His scientific interest is focused on modeling and performance evaluation of computer systems and computer networks,

combat modeling and simulation, mathematical methods of decision support. He has taken part in many scientific

projects connected with combat simulation and crisis management.

e-mail: rantkiewicz@wat.edu.pl
 Faculty of Cybernetics
 Military University of Technology
 Gen. S. Kaliskiego st 2
 00-908 Warsaw, Poland



Mariusz Chmielewski got his M.Sc. after individual studying in computer simulation as the 1st place graduate at Cybernetics Faculty Military University of Technology, Warsaw, Poland. He opened in 2007 his doctoral dissertation "Indirected associations method in semantic networks for crisis situation prediction". Since 2003 he has

worked as a lecturer at Cybernetics Faculty specializing in computer simulation and decision support systems. He participated in several projects. He is a member of NATO MSG026 project for Early Warning Systems for terrorist threat assessment.

e-mail: mchmielewski@wat.edu.pl
 Faculty of Cybernetics
 Military University of Technology
 Gen. S. Kaliskiego st 2
 00-908 Warsaw, Poland



Rafał Kasprzyk was born in Starachowice, Poland, in 1980. He received the B.Sc. and M.Sc. degrees in information science from the Faculty of Cybernetics, Military University of Technology, Poland, in 2005. He is currently working towards a Ph.D. degree. His main interest are game and graph theory, decision support system, computer simulation and homeland security.

e-mail: rkasprzyk@wat.edu.pl
 Institute of Computation Engineering
 Faculty of Cybernetics
 Military University of Technology
 Gen. S. Kaliskiego st 2
 00-908 Warsaw, Poland