ORGANIZING OF KNOWLEDGE BASES FOR EMERGENCY SITUATIONS USING SOCIAL NETWORKS

Oleksandr Kuzomin, Bohdan Tkachenko, Mariia Lysenko

Abstract: This paper describes technology of automated survey system for emergencies eyewitness that helps to collect data about emergencies.

Keywords: Social Networks, Emergencies Prevention, Automatic Survey, Language Analysis, Special Language.

Introduction

Procuring the security of human life, which is exposed to the risk of the natural disasters or environment changes, which may cause catastrophes (such as floods, flooding coastal zone of the Atlantic ocean, landslides, avalanches, fires, etc.) requires improving quality and amount of the information provision in Situation Centers [Kuzemin, 2004a].

At the moment there are many various services, which have their own databases with the information about known emergencies. But even specialized services not always have full and almost accurate information. That's why there is need in additional sources or ways to get new of more complete information about emergencies.

These problems can be solved with the help of systematic expansion of Databases through intelligent surveys, which automatically conduct a survey in social networks. Such surveys can collect information about emergencies that have already happened, about the assumptions that have caused them or immediately before emergency, and also about success of disaster management.

In the last years the fundamentally new civil society actors in the form of local structures - communities, or Online Communities began to play active role in prevention and elimination of the emergencies. Online Communities are relatively unstable group of people, interacting via the Internet-communications, provided by Internet services and having common interests and carrying out joint activities in virtual space [Parinov, 2000]. Habbermas thinks that the main characteristics of a network of civil society are "openness" as "establishment of broad, multi-dimensional relationships" and also "spontaneity" – "free formation, flow, and constant change" structures [McConnell, 2007].

Social networks are developed on a global scale. [Rimskiy, 2009] Social resources of online communities allow them to implement a number of functions by which the actualization of online communities as subjects of action in emergencies happens. According to a sociological study of the American Red Cross, dedicated to the role of social networks in information work in emergencies, 42% of respondents which received information from Internet, 6% of them already received information about emergencies from Twitter, 14%. In March 18 1999, American programmer Brad Fitzpatrick opened service of online blogging "Live Journal". Initially, project was planned as blog platform for communicating for friends of founder, but later grew into one of the most popular sites in the world. Today Live Journal is positioned as social network. In March 11, 2010 on site it is recorded more than 25 million registered users (in 2005 it was 8.9 millions, in June 12007 – 13 million).

For empowering situational centers [Kuzemin, 2004b] in predicting and disaster management, can play role above-mentioned social networks, which besides developing volunteer tasks for providing information support can be used to expand the database and knowledge about the processes in the emergency area. Herewith, is possible to get a versatile amount of system knowledge, which, using that knowledge precedent will reduce the time to search for alternative solutions and the adoption of effective measures for disaster management.

Proposed solutions

Greater involvement of social networks in choosing effective control decisions can be obtained by organizing a wide valid survey with the help of Social Networks. This survey should include structured sequence of questions about solving problem.

In the result of survey in Social Networks can be identified qualitative parameters, which represent a set of alternative decisions or actions. Every decision that a decision-chooser person is offered (DCP) has three main characteristics: purpose of the action; a description of the action; a means of its implementation. All these characteristics can be obtained from structured responses from Social Networks which can answer the following three questions: "Why?" or "What should be the result?" - The answer will characterize the goal (purpose of action); "What should I do?" - Answer will give description of the action; "How to do?" - The answer will be to characterize the means for performing (the ability and opportunity).

Foundation for building a "decision tree" based on system issues to identify knowledge for actions aimed for eliminating or preventing disaster has the following basic short questions: "What?”, "Why?”, "How?". Answers to these questions will create a knowledge base of the situational center.

Developing situation-linguistic model of the environment in the emergency area is aimed for creating production-framing hierarchy. Architecture of production-framing hierarchy is based on a frame-based knowledge, which will be based on framing inheritance hierarchy with attitude and active slots, and the output is production rules. This approach allows a natural way to combine in one model the statistical knowledge about emergency values in the form of slots, structural domain knowledge in the form of an inheritance hierarchy.

Thus, the frame-based systems can be represented as:

$$W : S \to I,$$

where I is a multiplicity of frames, $S = \{S_i\}, i = \overline{1,n}$ — finite set of slots of type $\langle v, d, \{D_j\}\rangle$, including the current value of the slot $v = \langle v_1, v_2, \dots, v_l\rangle \in T$ and default value $d = \langle d_1, d_2, \dots, d_k\rangle \in T$ of procedures demons $\{D_j\}$.

Inheritance relationship: induced by slot with reserved name parent: $F: G \Leftrightarrow \|F\ (parent)\| = G$. Typical for frame-based systems operation of specification for frame like on the model, is implemented by implicit inclusion to the model the rule $F(parent) \longleftarrow match(F, G)$. When considering multiple inheritances, slot parent is supposed as list type, and $F: G \Leftrightarrow G \in \|F\ (parent)\|$.

Conclusion about the values of the output parameter (resulting frame - I_D) is performed on condition of clear input values (request frame - I_c, consisted of subset $v \times d$) has the following form:

$$L: \left\{ \begin{array}{l} IF\ \langle v_1^1, \dots, v_1^1;\ d_1^1, \dots, d_k^1\rangle\ THEN\ \begin{matrix} P \\ 1 \end{matrix} \\ IF\ \langle v_1^2, \dots, v_1^2;\ d_1^2, \dots, d_k^2\rangle\ THEN\ \begin{matrix} P \\ 2 \end{matrix} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ IF\ \langle v_1^n, \dots, v_1^n;\ d_1^n, \dots, d_k^n\rangle\ THEN\ \begin{matrix} P \\ n \end{matrix} \end{array} \right\} \quad (1)$$

The resulting frame I_p represents aggregate of v and d, belonged to different frames. Thus I_p is output of the production system, which is presented as follows:

$$\langle S_1^p, S_2^p, \dots, S_n^p \rangle \to I^p.$$

To obtain this goal need to solve the following problems:

- Implementation of the construction of predicate queries and their modification, which will be the formally logical unit description and study the updating and modification of data and knowledge bases;
- Determine the rules of inference based on data and knowledge bases.

Under the concept of a database will mean a set of facts that we get from social networks as a result of the survey. The main ideas of this approach are considered under specific SQL implementations or implementations of Internet technologies.

Description of the Knowledge Base

All decisions in the subject area are taken on the basis of analysis of the conclusions of experts and specialists with work experience. Information system Knowledge Base examined according to [Slipchenko, 2004] as set of information entities of atomic predicates from some information space $\mathfrak{R}$. All changes that occur in the knowledge base are considered as a consequence of the modification of predicate queries Q_m. The basis of these predicate queries is a set of modification predicate rules:

$$Q_m \leftrightarrow (K_B) \ll \left\| \begin{matrix} K_{B-(X)} \\ K_{B+(X)} \end{matrix} \right. ,$$

where $X \in \mathfrak{R}$.

$K_{B+(X)}$ means, that atomic predicate $^{\mathrm{O}}$ should be included in knowledge base K_B, K_B means, that X should be excluded from the knowledge base; $(K_B) \ll$ means modification of the knowledge base at the level of logical consistency of predicate rules; $K_{B\pm(X)}$ means the ability to modify not only the knowledge base, but also protection of the user based on the descriptors; $\ll$ considered as a complex arrow, features of are examined by the theory of categories.

Knowledge Extraction

Knowledge can be represented in the form of production rules of the type [3]:

if $X_1 \& \dots \& X_K$ *then* $X_{K+1} \& \dots \& X_{K+L}$,

where $X_1 \dots X_K$, $X_{K+1} \dots X_{K+L}$ - some predicates.

Definition 1. Content of knowledge if $X_1 \& \dots \& X_K$ *then* $X_{K+1} \& \dots \& X_{K+L}$ is called set $W = \Pi_1 \times \Pi_2 \times \dots \times \Pi_{K+L}$. Arbitrary element of this set is an element of content knowledge.

Content of condition knowledge is called a set $W_1 = \Pi_1 \times \Pi_2 \times \dots \times \Pi_K$. Arbitrary element of this set is called an element which contains conditions of knowledge.

Content of consequence knowledge is called a set $W_2 = \Pi_{K+1} \times \Pi_{K+2} \times \dots \times \Pi_L$. Arbitrary element of this set is called an element which contains consequence of knowledge.

Definition 2. Under the probability p_i of element the knowledge content $w_i \in W$ we will mean the probability of event, consisting in the fact that all predicate constants up the w_i will take a logical value "AND" by substituting the value of objects instead of arguments from areas of truth predicate variables that make up this knowledge.

Element w_j which contains knowledge is a vector, components of which are the values of predicate variables, included into the knowledge. We can associate a vector $z_j = z_j(1), \dots, z_j(K+L)$ from R^{K+L} with an element w_j which contains knowledge.

The function of the distribution of knowledge is a function from $K+L$ arguments: $F(y) = F(y(1), y(2), \dots, y(K+L))$ with domain of definition R^{K+L} and taking values in the space R^1. It is defined by the formula $F(y) = \sum_{z_j \le y} p_j$, where z_j displaying the element of the content of knowledge w_j in R^{K+L}. $z_j < y$ expression is understood as an act of: $z_j(i) < y(i), i = 1, \dots, K+L$.

Definition 3. As distance between comparable knowledge 3H1 and 3H2 we will call Hellinger distance $d(G,Q)$ between two probability distributions of their content elements $G = \{p_{11}, p_{12}, \dots, p_{1r}, \dots\}$ and $Q = \{p_{21}, p_{22}, \dots, p_{2r}, \dots\}$, which is calculated by the following formula:

$$d(G,Q) = \sum_j (\sqrt{p_{1j}} - \sqrt{p_{2j}})^2 \quad (2)$$

Calculating the distance between knowledge, can be solved the problem when the input of the knowledge base is supplied some new knowledge. Required to determine which of the subsets of the knowledge should include this new knowledge. To solve this problem we should calculate the distance between the new knowledge and all available knowledge, and then take it to a subset containing such knowledge, for which the distance is minimum.

Decision making

Despite the fact that decision-making is carried out in a selected subset of knowledge for complex systems and processes adequate mathematical description of the decision is absent or are rather cumbersome mathematical constructions which optimization and practical use in real-time is impossible. This problem can be solved, using algorithms, based on models, simulating the decision-making process by experienced expert [Kuzemin, 2005]. For a large number of models decision-making as the mathematical apparatus can be used the theory of fuzzy sets. When choosing solutions in situational centers the aim is to choose the design stages of design options or parameter value from a fairly small predetermined set determined as mentioned earlier, using the formula (2). For modeling decision-making process proposed to use decision-making models, based on fuzzy rule modus ponens, inductive output fuzzy and fuzzy expert information of the second kind. This will be used Inductive Output type [4]:

$$L': \left\{ \begin{matrix} IF\ B'_1\ THEN\ A'_1 \\ IF\ B'_2\ THEN\ A'_2 \\ \cdots\cdots\cdots\cdots \\ IF\ B'_m\ THEN\ A'_m \end{matrix} \right\} \quad (3)$$

Clear-cut statements A' and B' have the form:

$$A': \langle \beta_w\ is\ w' \rangle;\ B': \langle \beta_v\ is\ v' \rangle;$$

$$w' = (x, y, z, \dots) \in X \times Y \times Z \times \dots, v' \in V$$

In this scheme of output statements about the values of the input parameters are sending to schema itself (statement A') and a consequence inside the system L' of the statement (statement A_j) and statements about the values of output parameters are the consequence of output circuits (3) (statement B'), but sending circuit inside L' (statement B'_j). Therefore, to select the output v parameter values, based on the rules modus ponens it should convert the output circuit (3) in the form:

$$\left.\begin{matrix} L' \\ A' - true \end{matrix}\right\} \rightarrow B' \ is \ true$$

For this purpose is proposed to convert system of the statements of the second type in the equivalent system of the first type, using the contra positive rule, according to which for arbitrary expressions A and B saying "$IF\ A\ THEN\ B$" and "$IF\ \bar{A}\ THEN\ \bar{B}$" are equivalent.

Applying the rule to the expressions $L'_j, j = 1, \dots, m$ contrapositive of the second type, we obtain

$$\langle IF\ A'_j\ THEN\ B'_j \rangle = \langle IF\ \overline{A'_j}\ THEN\ \overline{B'_j} \rangle$$

where statements $\overline{A'_j}$ and $\overline{B'_j}$ can be considered as statements of $\langle \beta_w\ is\ \alpha_k \rangle$ and $\langle \beta_v\ is\ \alpha_m \rangle$ in which the values $\alpha_{w_j^*}$ and $\alpha_{v_j^*}$ are determined by membership functions $\mu_{w_j^*}$ and $\mu_{v_j^*}$ which are in addition to $\mu_{w_j}$and μ_{v_j}:

$$\mu_{w_j^*}(w) = 1 - \mu_{w_j}(w), \forall w \in W = X \times Y \times Z \times \dots;\ \mu_{v_j^*}(v), \forall v \in V$$

Knowledge base construction

Emergency description $Sit' = \{Sit_i\}\ (i = \overline{1,n})$ consists from the set of micro situations – s_i, which are formed by concepts – environmental elements that are being presented by survey participants:

$$Sit' = \langle e_i, K_e, X, \mathfrak{I}, \Sigma \rangle$$

where Σ – required resources (volumes and methods for the prevention and elimination of emergency situations); $\mathfrak{I}$– a plurality of control actions; part of the situation that is defined by pair $\langle e_i, K_e \rangle$ is called linguistic (qualitative, semantic unit) micro situation of central concept – e, which is based around micro situations.

$X = \{x_j, (j = \overline{1,m})\}$ – quantitative indicators about natural phenomena (e.g., atmospheric pressure, temperature, etc.), constant parameters of the environment (for example, the slope of the mountain, covering the slope, etc.) in a controlled area and resources that respondent was able to identify (e.g., the number of serviceable mobile means for evacuation, mobile facilities for aerial reconnaissance, etc.). Set $K_e = \{Ot_i, Pvt_k\}$ is geo-information context of the micro situation for linguistic, central concept e. Set $K_c = \{k_{e_i}\}, i = \overline{1,m}$ expresses relationship Ot_i of the central concept e with other minor concepts Pvt_k that are being participating in current micro situation. Relationship Ot_i is some dependency of the central concept from minor concepts. For current micro situation as the minor concept can act concept that is a central concept to another micro situations.

For the text, obtained in the survey through social networks - must be extracted from the set of candidates to the central concepts $cPvt$ that will be associated with a variety of micro situations Sit' precedents that represent knowledge about the previously observed emergency or disaster that was observed by survey participant.

Stages of building standard or observed micro situations:

1. Description of the current situation of the domain in the form of narrative text;
2. Identification of concepts obtained from the descriptions from the categories;
3. Search relations between these concepts;
4. Building the description in a language of representation of micro situations.

Concept Pvt for considered category $eKat$ receive a result of executing the function to identify the concepts ERp

$$ERp(TextExp) \rightarrow Pvt_i$$

where Pvt_i – identified concepts, $i = \overline{1,n}$.

And we have

$$Sit' \Longrightarrow \{cuP, Ot\}$$

where Sit' micro situation, for which are corresponded: cuP – a set of concepts, which are entities and Ot – a set of concepts that express the relationship between the other concepts.

Step 1. Choose from a variety of candidates cuP set of central concepts or precedents - $cPvt$ (nouns that are the subject of a sentence $TextExp$).

$$ERp1(cuP) \rightarrow cPvt_i$$

where $ERp1$ is a function of identifying candidates into central concepts - $cPvt$

Step 2. Choose context from relations.

$$OtA = {}^{cuP}/_{\tilde{n}Pvt}$$

for candidates received in the central concepts $cPvt$. Task is to extract a subset of relations (associations) $Ot \subset OtA$.

Elements of the set of relations Ot are the main, active and additional. Each of the set of relations $OtP_1 \in Ot$ will have relation to a central concept $cPvt_i \in cPvt$.

Step 3. Forming set of micro situations $Sit'_i = \langle cPvt_i, Ot_{cPvt_i} \rangle$ - precedents, in which elements of the sets Ot_{cPvt_i} are elements of the set Ot. At this stage received micro situations still are not complete as to the elements of Ot not yet mapped to the secondary concepts.

Step 4. Search secondary concepts. As secondary concepts are any elements of the set of candidates cuP regardless of whether they were in the set of central concepts $cPvt$, set of relations Ot or any one of them. In most cases, secondary concepts derived from additions to verbal forms, previously selected. In general, secondary concepts - are the concepts that are referenced by the association.

The decision rule is represented as follows:

$$pravResh = \underset{i}{\wedge} Prizn_i$$

where $Prizn_i$ – separate characteristic.

When calculating the value of a decision rule $pravResh$ each of signs $Prizn_i$ is assigned a value of true if this feature is present in the concepts and false otherwise. Denote a text description of the situation Sit obtained from an expert as $TextExp$. The notion of a primary domain text representation $etExp$ a concept expressed in terms of existing categories. The process of identifying concepts Pvt from the $TextExp$ is a function that displays text on a set of concepts:

$$ERp(TextExp) \rightarrow Pvt_i$$

where ERp– function to identify the concept; Pvt_i – identified concepts, $i = \overline{1,n}$.

For micro situation Sit' we have two sets of candidates: cuP – a set of concepts, which are entities and Ot – a set of concepts that express the relationship between the other concepts.

Conclusion

The proposed approach is the basis for establishing an expanded knowledge base situation center for forecasting and disaster management. Using the model representation and technology knowledge base allow you to create precedent knowledge base for decision-making for considerably less time using extensive interviews eyewitnesses disaster through social networks.

Acknowledgement

The paper is published with the financial support from the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA (www.ithea.org) and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine (www.aduis.com.ua).

Bibliography

[Kuzemin, 2004a] Kuzemin A. Situation centres in modern sate // International Journal on Information Theories&Applicatios.– Bulgaria, 2004. – Vol. 11, №1.– P. 79 - 82

[Kuzemin, 2004b] Kuzemin A., I. Yanchevskiy, Sorochan M., Torojev A. The use of situation representation when searching for solution in computer aided design systems // International Journal on Information Theories&Applicatios.– Bulgaria, 2004. – Vol. 11, №1.– P. 82 - 88.

[Kuzemin, 2005] Kuzemin A., Sorochan M., Yanchevskiy I., Torojev A. The use of situation representation when searching for solutions in computer aided design systems International Journal // Information Theories & Applications. – Bulgaria. – 2005. – Vol.11. – P. 101 – 107.

[McConnell, 2007] McConnell B., Huba J. Citizen Marketers: When People Are the Message

[Parinov, 2000] Онлайновые сообщества: методы исследования и практическое конструирование. http://rvles.ieie.nsc.ru/~parinov/autoref.htm (accessed: 31.03.2014)

[Rimskiy, 2009] Римский В. Воздействие сети Интернет на социальную активность, формирование и развитие идентичностей // Вестник общественного мнения. 2009. № 1.

[Slipchenko, 2004] Slipchenko M., Kuzemin A., Sorochan M., Torojev A Situation centers in modern state // Proc. of the Second International Conference i.TECH. – Sofia. – Bulgaria: ITHEA. – 2004. – P. 96-98.

Authors' Information

Oleksandr Kuzomin – Prof. Dr., Department of Informatics, Office of Innovations & Marketing, Kharkiv National University of Radio Electronics; Kharkiv, Ukraine; tel.: +38(057)7021515; e-mail: kuzy@daad-alumni.de

Major Fields of Scientific Research: General theoretical information research, Decision Making, Emergency Prevention, Data Mining, Business Informatics.

Bohdan Tkachenko – Master of Kharkiv National University of Radio Electronics; Kharkiv; Ukraine; tel.: +380 97 238 61 31; e-mail: bogus.weber@gmail.com

Major Fields of Scientific Research: General theoretical information research, Knowledge Discovery and Engineering, Business Informatics.

Mariia Lysenko – Master of Kharkiv National University of Radio Electronics; Kharkiv; Ukraine; tel.: +380 68 349 29 09; e-mail: trendskaut@gmail.com

Major Fields of Scientific Research: General theoretical information research, Knowledge Discovery and Engineering, Business Informatics.