# Risk Modeling During Complex Dynamic System Evolution Using Abstract Event Network Model

Volodymyr Sherstjuk<sup>1</sup>, Maryna Zharikova<sup>2</sup>, Ibraim Didmanidze<sup>3</sup>, Irina Dorovskaja<sup>1</sup> and Svitlana Vyshemyrska<sup>1</sup>

#### Abstract

The paper is devoted to the issues of knowledge representation about a plurality of processes of various intensities that arise unexpectedly and evolve simultaneously in complex dynamical systems in a wide range of domains. A knowledge representation model based on events that are referenced to certain time intervals and spatial areas is proposed. Since some information about events is inaccurate or blurred, the uncertainty of the observed information about the evolving processes is represented using gray numbers and soft sets. Events are connected in an abstract network, where arcs express possible transitions from one event to another. Within the network, transitions between events are driven by impulses and correspond to transitions of a dynamic system from state to state. The energy accumulated in the nodes is considered to generate impulses, which express the achievement of a certain threshold by the energy. The proposed abstract event network can be used to model the evolution of dynamic systems that can be expressed by events that occurred inside the system but driven by impacts from the outside (environmental effects). Such evolution is considered in a wide range of rates, from the slowest processes associated with climate change to the most rapid processes associated with the effects of natural forces of a destructive nature. Connections between nodes allow representing sequential and parallel streams of events concerning cascade and triggering effects, which makes it possible to study complex interactions between separate processes within a complex system of a random structure.

#### **Keywords**

knowledge representation model, dynamic system evolution, energy-driven abstract event network, uncertainty model, temporal and spatial referencing

#### 1. Introduction

There are many domains of technical, socio-technical, or socio-economical nature that evolve in space and time within the natural environment. Usually, they can be represented by complex

CITRisk'2021: 2nd International Workshop on Computational & Information Technologies for Risk-Informed Systems, September 16–17, 2021, Kherson, Ukraine

ORCID: 0000-0002-9096-2582 (V.Sherstjuk); 0000-0001-6144-480X (M.Zharikova); 0000-0001-6695-4980 (I.Didmanidze); 0000-0001-9280-8098 (I.Dorovskaja); 0000-0002-6343-7512 (S.Vyshemyrska)



© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<sup>&</sup>lt;sup>1</sup>Kherson National Technical University, Beryslavske shose 24, Kherson, 73008, Ukraine

<sup>&</sup>lt;sup>2</sup>Universität der Bundeswehr, Werner-Heisenberg-Weg 393GEB, München, Neubiberg, 85579, Germany

<sup>&</sup>lt;sup>3</sup>Batumi Shota Rustaveli State University, Ninoshvili/Rustaveli str. 35/32, Batumi, 6010, Georgia

EMALL: vgsherstyuk@gmail.com (V.Sherstjuk); maryna.zharikova@unibw.de (M.Zharikova); ibraim.didmanidze@bsu.edu.ge (I.Didmanidze); irina.dora07@gmail.com (I.Dorovskaja); vyshemyrska.svitlana@kntu.net.ua (S.Vyshemyrska)

dynamic systems containing multitudes of interacting dynamic objects. Due to such interactions, a plurality of processes of various intensity arises unexpectedly and evolve simultaneously within dynamic systems [1]. Typically, these processes are transient, non-linear, and non-stationary. The dynamic objects have specific states at specific times, which are represented by certain points within the state space of the dynamic system, so their behavior is usually represented by their trajectories within this space. Clearly, if their trajectories touch or intersect each other (and even in the case of some proximity of the trajectories within the state space), most often these objects are exposed to a variety of dangers, threats, and risks.

Some processes such as climatic changes proceed very slowly, others, on the contrary, very quickly, so they can be dangerous and cause deaths, injuries, and huge damage [2]. Since the processes, which evolve and interact within complex dynamic systems, are principally stochastic, the dynamic system has a random structure and is difficult to control. The evolution of the dynamic systems is mainly driven by the impacts of humans and nature [3]. The manifestation of such evolution can be ex-pressed by events that occurred inside the dynamic system but associated with im-pacts from the outside (environmental effects). Due to the wide dispersion of a plural-ity of the evolving processes over space and time, events should be properly referenced within space and time.

At the same time, people must be concerned about the prevention or elimination of dangerous events to minimize losses, so there is a need to ensure the safety of the dynamic system [4]. To make the dynamic system safe, people need first to observe the system itself and the environment. Although there are a lot of technical means of observation (i. e. sensor networks, unmanned vehicles, etc.), all of them are based on sensors, which usually provide ambiguous, imprecise, incomplete, inconsistent, and doubtful information. Such a wide range of uncertainties distorts observations and introduces unpredictability of the states of the dynamic objects and the dynamic system as a whole. Naturally, this requires the development of special safety-enabled methods to overcome uncertainty both in the spatial and temporal aspects [5].

Second, since a human cannot directly control the flow of processes within the complex dynamic system, he must be able to make decisions, the implementation of which allows him to change the flow of processes indirectly. Thus, people need to use decision support systems [6]. However, due to the above-mentioned reasons, decision support is a complex and non-trivial task.

The spatial aspects of the dynamic system can be typically expressed by a geo-graphic information system (GIS), which contains a spatial model and outlines an area of interest (AOI) critical for decision making. Thus, the events distributed over the AOI can be properly referred to spatially, but their references are usually limited by measurement inaccuracy.

The temporal aspects are more complicated due to their stochastic nature and simultaneous impacts. Moreover, it is very difficult to refer to events temporally be-cause of the different rates of their occurrence. Unlike spatial references, which usual-ly are vague or blur due to the sensors' imprecision, temporal references are inaccurate and can be primarily represented by time intervals.

Since decision-makers usually operate under the conditions of high responsibility and a lack of time, the decision support system must work in real-time. Thus, considering the above reasons, the most topical and important issue for today is the development of GIS-based real-time decision-support systems (DSS) aimed at danger and risk assessment for the considered class of complex dynamic systems.

Unfortunately, today there is a lack of knowledge representation models and methods that would be operable in real-time and take into account a wide range of measurement uncertainties

in data captured by sensors during the observation of the dynamic system. Thus, the research of knowledge representation methods based on events and adequate uncertainty models that allow describing dynamic systems by spatially and temporally referred trajectories is a topic of our interest. The problem addressed in this paper relates to the representation of knowledge about dynamic system evolution using abstract event networks from the point of view of developing real-time DSS.

#### 2. Recent Works

Representation of uncertain knowledge about events that occur simultaneously and jointly has been studied in many fields of knowledge. A wide range of models of knowledge representation about events has been proposed in the field of natural language processing [7]. Mainly, such models are focused on event-to-event relations and processing of events to discover their causal relationships and detect anomalies within language structures [8].

In existing risk assessment decision support systems, the most frequently observed information is represented by event streams, which describe sequences of time-referred events [9]. Mainly, the event stream model is a subject of study in event sequence analysis, which focuses on the time gaps between events and their order in the sequence [10]. Another trend is the use of knowledge structures based on events as certain building blocks to build and update situation models [11]. Thus, a most used definition of an event has been proposed in the literature as "a segment of time at a given location that is conceived by an observer to have a beginning and an end" [12].

The above-mentioned approaches are mainly semantic based, so they use strict definitions of events, time, and space and take little into account the uncertainty of the information, based on which the events must be determined [13].

Many non-semantic methods have been proposed including Causal events, Force Dynamics, Stochastic Context-Free Grammars [14] that represent complex event structures based on hierarchical definitions, which are hard for decision-maker interpretation and do not meet the requirements for real-time DSSs.

In general, considering the nature of the events, we emphasize two main approaches to represent events - probabilistic and non-probabilistic. The first one includes such methods as Hidden Markov Models, dynamic Bayesian networks, Monte Carlo sampling, Variance propagation, etc. [15]. The second one includes methods mainly based on fuzzy sets and possibility theory [16]. Obviously, a lack of sufficient statistical data for probabilistic approach as well as a lack of well-known possibility degrees or membership functions for non-probabilistic approach complicates their efficient use and leads to high computational complexity. Detailed overviews of such approaches concerning a considered class of DSSs have been presented in [17].

The event tree approach enables the modeling of a sequence of events, which constitute the structures of any level of complexity adapted to various uncertainty models (probabilistic, fuzzy, rough, etc.) [18]. Despite the flexibility of this approach and its potential for evolution, the existing event tree models refer events to time only, so there is a lack of spatial localization of events. A method for representing hierarchical structures of events referred both to time and space and equipped with a complex uncertainty model has been proposed in [19]. However, the computational complexity of the proposed method is significant, which complicates its use in real-time DSSs.

Thus, we conclude that existing event-based knowledge representation models can be weakly applied to DSSs of the considered class. We need not only to refer events to a certain time and spatial locations but also use a relatively simple but effective model for the representation of uncertainty that could satisfy both the requirements for the representation of incomplete and inaccurate information obtained from the observation of a dynamic system and requirements to efficiency that makes a model suitable for real-time GIS-based DSS.

## 3. Uncertainty Models

Last years, researchers have directed sufficient efforts towards improving the above-mentioned issues. The classical approach [20] is that a behavior of a dynamic system can be represented by high-dimensional nonlinear equations, which describe complex processes arising within the considered space and time. However, such continuous modeling of dynamic systems contradicts a lack of data of required quality and accuracy caused by uncertainty and imprecision of sensor data as well as its discontinuous measurements. Furthermore, systems of high-dimensional equations cannot be solved in a reasonable time due to a significant computational complexity [21]. The statistical approaches [22] also cannot provide justified assessments primarily be-cause of a lack of reliable statistics and weak observability of a dynamic system.

Since the use of the continuous space, time, and correspondent models of the dynamic system leads to a huge computational complexity, it is advisable to discretize time and space (e.g. AOI) [23] for better compatibility with discrete measurements provided by sensors and to speed up the calculations. At the same time, the use of a discrete model reduces the accuracy and credibility of assessments [24]. Therefore, based on the domain features, it is necessary to choose such a sampling discrete that provide both the sufficient performance of assessments and the required accuracy of its results.

The main question is how to manage uncertainties. The most commonly used approach to take into account uncertainties is probabilistic, which represents different aspects of uncertainty in terms of chances [25]. However, typically there are few repeated occurrences of events under the same conditions, especially considering their spatial and temporal references. Thus, the probabilistic approach deals only with stochastically stable data and represents uncertainty inadequately [26].

In this regard, various non-probabilistic methods of uncertainty modeling have been developed such as fuzzy, rough, vague, soft, grey sets, etc.

Obviously, real data cannot be represented as crisp and well-determined. Instead, sensors provide data that can be not clearly known, undetermined, problematic, varying, vague, can have many interpretations but not certain information, and, of course, can be no reliable [27]. Discretization of space and time will also lead to various sampling errors, delays in time, outdated data, etc. [28].

Thus, there are a lot of types of uncertainties that should be modeled and processed within event-based knowledge representation about complex dynamic systems. Let us consider the available tools for modeling uncertainty.

Zadeh introduced the concept of the fuzzy set [29] that has been used in a wide range of various fields and proposes a convenient tool to represent vague data. Its drawbacks are that fuzzy sets are computationally hard, and their membership functions are subjective and can be difficult to found [30].

The long-term study of many researchers has led to the emergence of dozens of fuzzy extensions and additions both in theoretical and practical areas [31], however, they do not allow

to overcome the above disadvantages, and further increase the computational complexity, which hinders their use.

Pawlak introduced the concept of rough sets [32] as a mathematical tool to deal with imprecise or noisy data based on equivalence (e.g., indiscernibility) relation, that fits well to discrete-valued and nominal data. Rough sets are easy to understand, suitable for inconsistent data, and do not need any additional information about data. Their drawbacks are a problem with dirty or noisy data, depending on complete information, and a lack of membership values. In general, in many fields of application, the rough set algorithms are much more efficient than fuzzy.

Clearly, fuzzy sets and rough sets provide models for two different types of uncertainties. Fuzzy sets can be represented by their membership values, but rough sets can be approximated through partitions. Fuzzy sets highlight the vagueness of information while rough sets focus on incomplete information. Therefore, fuzzy sets mainly represent subjective uncertainties while rough sets represent objective uncertainties.

Grey sets and grey numbers have been proposed by Deng Julong [33]. A grey number is a number, whose exact value is not known but its interval is known. In general, the grey number introduces a certain set of values and represents only one number, which is not clearly identified among the elements of the set. It can be reduced to a white number or black number, the first one is an exact or crisp value while the second is a number, whose exact value or value's interval is not known. Thus, grey numbers can be an adequate model in the case when the exact value is not completely known.

Molodtsov introduced the concept of the soft sets [34] as a relatively new approach to model both uncertainties and vagueness, which is free from the difficulties of existing methods. Soft set membership can be determined through a certain parametrization given by real numbers, functions, mappings, and others, even words or sentences. Thus, a membership function problem cannot arise in the soft sets. They are a convenient and easy tool for model both objective and subjective uncertainty, their main advantage is that they are free from the inadequacy of parameterization tool. The drawback is that the soft set does not assign any membership values. Soft sets are rarely used independently, but often become the basis for complex models for representing uncertainty.

Obviously, in the real world, objective uncertainty and subjective uncertainty may exist at the same time. Accordingly, researchers try to build complex models that consider different types of uncertainty at once either by extending existing models or by combining them. Often, the uncertainty models presented above are related and complementary to each other. Moreover, they can also be reduced to one another [35]. This makes it possible to combine them, obtaining complex models of uncertainty.

The most effective uncertainty models are soft and gray. In this work, we propose to use gray numbers to represent the values of the observed parameters of the dynamic system as well as soft sets to represent events concerning their classes, temporal and spatial locations that make them applicable to the analysis of dynamic processes.

Let us consider a model of a dynamic process at two levels: at the micro level considering individual spatial elements, and at the macro-level considering spatial areas of a larger or smaller scale.

## 4. Micro-Model of Dynamic Processes

### 4.1. Cells and their states

Consider an AOI as a two-dimensional Euclidean space discretized uniformly by a metrical grid D. Using the spatial discrete  $\delta$ , the grid D outlines the two-dimensional array  $D = \left\{d_{xy}\right\}_{x,y=0}^{N}$  of square cells  $d_{xy}$  of size  $\delta \times \delta$ , where x and y are the array indices corresponding to the coordinate axes. Suppose a certain cell  $d_{xy} \in D$  is an object of consideration that represents a minimal homogeneous area within the AOI.

Let us imagine that a certain non-empty set of parameters  $A = \{a_i\}_{i=1}^m$  can be obtained as a result of the observation and associated with the cell  $d_{xy} \in D$ . Let  $V_{a_i}$  be a domain of each  $a_i \in A$ ,  $V = \bigcup_{a_i \in A} V_{a_i}$ , and f be a value function  $f : A \times D \to V$  that returns a value of a certain parameter for the cell  $d_{xy} \in D$ .

If the values of some parameters  $a_i \in A$  are unchanged over time, such parameters belong to a subset  $A_S$  of static parameters,  $A_S \subseteq A$ . In this way, if the values of certain parameters are varying over time, such parameters constitute a subset  $A_D$  of dynamic parameters,  $A_D \subseteq A$ . There can also be such parameters whose values change over time, but slow enough (e.g., environmental parameters), they constitute a subset  $A_E$  of slowly changing parameters. Obviously,  $A = A_S \cup A_D \cup A_E$ .

Each cell  $d_{xy} \in D$  can be associated with a certain subset  $A(d_{xy}, t) \subseteq A$  of parameters' values at a certain time t; some of them can be imprecise while others can be unobservable at the time t. It should also be noted that some of the parameters cannot be directly measured or estimated, so they require to use of indirect methods.

Let us consider a state  $S(d_{xy},t)$  of the cell  $d_{xy}$  at the time t such that  $S(d_{xy},t) = \{a_m(t)\}_{m=1}^z$ ,  $\forall a_m(t) \in A(d_{xy},t)$ , and a state function  $v: D \times A \to S$ , which returns the state of the cell  $d_{xy}$  diagnosed at the time t based on the observed subset of the cell's parameters.

Let  $C = \{c_j\}_{j=1}^q$  be a set of cell's statuses and  $\mathcal{G}$  be a status function  $\mathcal{G}: D \times A_s \to C$ . Thus, we can correlate a cell with a particular status based on information about the values of a certain subset of its static parameters. For example, a cell can have a status of "water", "soil", "sand", or "rocky" based on the value of the observed parameter "ground". Obviously, the meaning of the cell's status can vary but it should be based on a selected set of static parameters. Similarly, the state of the cell can be defined with respect to the different points of view.

Let  $W = \{w_k\}_{k=1}^n$  be a set of state classes. Suppose each state class  $w_k \in W$  contains a finite set of micro-states,  $w_k = \{w_{kl}\}_{l=1}^f$ , which constitute an ordered sequence of transitions  $[w_{kl}, ..., w_{kf}]$ , where  $w_{kl}$  is an initial micro-state and  $w_{kf}$  is a final micro-state. Concerning the dynamic process, we assume that the dynamic process covers the cell when the last being in the micro-state  $w_{kl}$  and ends within the cell when the cell enters the state  $w_{kf}$ . Obviously, each micro-state  $w_{kl}$  can be defined as a certain subset of the cell's state at the time t determined by a specific subset of parameters  $A_{kl} = \{a_{klm}\}_{m=1}^u \subseteq A$ .

Fig. 1 shows the example of two classes of states, namely "moisture" and "burning", defined by sequences of micro-states  $w_1$  and  $w_2$ . Some microstates can be observed simultaneously, for example, the cell can be defined by micro-states "dry" and "pre-ignited" at the same time. However, micro-states "wet" and "pre-ignited" cannot coexist. Another peculiarity of the proposed micro-model is that micro-states of a certain class are not always compatible with every possible cell status. For example, the cell statuses "water" or "rocky" is not compatible with the state class "moistening" for obvious reason. Clearly, we need to define a state class function  $\omega: D \times A \to W$  that returns a micro-state of the cell based on the values of its parameters that belong to  $A_M$ .

We assume that each transition of the cell  $d_{xy}$  from one micro-state  $w_{kl}$  to another micro-state  $w_{km}$  is a micro-event  $\psi: w_{kl} \to w_{km}$  (Fig. 2).

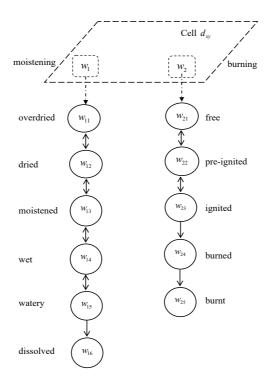


Figure 1: Representation of micro-states of the cell

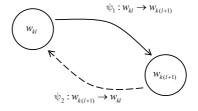


Figure 2: Example of direct and reverse transition of a cell from one micro-state to another

Thus, the dynamic process can be modeled by sequences of dynamic changes of micro-states of cells covered by the process spatially. During the process, cells pass through a sequence of micro-states. Obviously, some transitions from one micro-state to another in the context of a certain state class can entail transitions within other state classes. Moreover, the transitions of micro-states of one cell can spread to neighboring cells, which makes it possible to simulate the propagation of the dynamic processes within AOI.

## 4.2. Propagation of the Dynamic Processes

We assume that all changes in the micro-states of cells, as well as all conditions for the propagation of a dynamic process between cells, are associated with the transfer of certain contingent energy.

Suppose that each class of states has its class of energy. Energy can be generated because of changes in the values of the cell parameters and accumulated in the specific energy storage inside the cell.

There are a set I of energy transfer channels to transfer energy between cells as shown in Fig. 3. Eight channels  $\{I_{+,+},I_{+,0},I_{+,-},I_{-,0},I_{0,+},I_{-,-},I_{0,-},I_{-,+}\} \in I$  reflect the relative position of cells in space and therefore are denoted relative to the considered cell.

These channels can be unidirectional or bidirectional.

Since the propagation of the process is usually influenced by a significant number of factors of a stochastic nature located outside the system under consideration, it is necessary to have a specific tool for modeling such effects. Many dynamic systems are influenced by such external processes associated with the environment. For example, the speed and direction of propagation of a forest fire are most influenced by the speed and direction of the wind. Clearly, the wind is a separate dynamic process observable by dynamic parameters, which should be taken into account in the model.

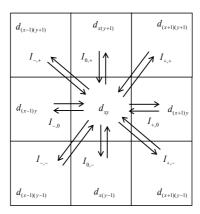


Figure 3: Channels for energy transfer between cells

Thus, it is proposed to use a matrix  $\Omega$  of coefficients of energy transfer through channels, where each coefficient can slow down or accelerate the process of energy transfer, so the speed of the dynamic process propagation can change accordingly.

The matrix  $\Omega$  can be defined as

$$\Omega = \begin{bmatrix} \alpha_{+,-} & \alpha_{+,0} & \alpha_{+,+} \\ \alpha_{-,0} & 0 & \alpha_{+,0} \\ \alpha_{-,-} & \alpha_{0,-} & \alpha_{+,-} \end{bmatrix},$$

where each coefficient  $\alpha_{*,*}$  corresponds to a certain energy transfer channel  $I_{*,*}$  and takes values in the range [-1,1]. Direct determination of the values of the coefficient matrix  $\Omega$  is beyond the scope of this paper, but its source is obviously the observations.

#### 4.3. Micro-Events

In this work, each micro-event is referenced in space and time. The impulse paradigm is used to model micro-events, which are considered as the direct effects of the change of values of specific observable parameters.

A unique descriptor  $y_i$  can be defined as a tuple:

$$y_{j} = \langle a_{j}, \lambda_{j}, \delta_{j}, \Delta_{j} \rangle, \tag{1}$$

where  $a_j$  is a certain parameter,  $a_j \in A_{kl}$ ,  $\Delta_j$  is an absolute change of the value  $a_j$ ,  $\lambda_j$  is a susceptibility for the attribute  $a_j$ , and  $\delta_j$  is a threshold value.

Due to the inaccuracy of observations, the estimated values of parameters are mainly inaccurate. This requires adequate representation. To ensure correct representation, we can define intervals that contain the confidence degrees from the minimum to the maximum possible value. Since such intervals can be narrowed during further observations, we propose to represent the value of the observed parameter as a gray number. Moreover, it is also necessary to represent the threshold value as a gray number to achieve a trigger effect, which is often observed in real natural systems.

Thus, the value  $\Delta_j$  can be described as the interval  $\Delta_j^{\pm} = \left[\Delta_j^-, \Delta_j^+\right]$  that represents a grey number and  $\delta_j$  can be "grayed" in the same way. The descriptor  $y_j$  also turns gray and is denoted by  $\hat{y}_j$ . An integral descriptor y can be defined as a tuple:

$$y = \left\langle \left\{ \beta_j \cdot \widehat{y}_j \right\}_{j=1}^m \right\rangle,\,$$

where  $\beta_j$  is a certain coefficient for each corresponding descriptor  $\hat{y}_j$ . It represents the direct effect of the simultaneous change of several parameters' values within the exposed cell.

(2)

Consider a micro-event as a consequence of the transition of a cell's micro-state of a certain class, which can be generated as an energy impulse  $\varepsilon$  based on y.

The impulse generation scheme is shown in Fig. 4.

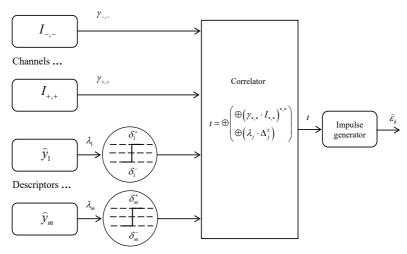


Figure 4: The impulse generation schemes

It considers the fact that the change in the values of the parameters is influenced by the impulses received through the channels from the neighboring cells. The correlator (Fig. 4) accumulates inputs and ensures that the control stimulus is found, and the impulse generator produces an energy impulse of an amount  $\varepsilon_k$  that depends on the magnitude of the control stimulus  $\iota$ . The descriptors' values are compared with a threshold value. The result of the descriptors' triggering helps us to determine whether an event will occur.

Thus, let us describe the micro-event  $\psi_k$  as a couple

$$\psi = \langle t_i, d_{xy}, w_k, w_{km}, \varepsilon_k \rangle, \qquad (3)$$

where  $t_j$  is a time reference of  $\psi$ ,  $d_{xy}$  is a spatial reference of  $\psi$  within AOI,  $w_{km}$  is a new  $d_{xy}$  micro-state of state class  $w_k$ , and  $\varepsilon_k$  is an amount of generated energy of class k that represents an event magnitude.

## 5. Macro-Model of Dynamic Processes

At the higher level, the grid D can be divided into disjoint objects, which describe the homogeneous areas of the AOI in terms of their parameters' values.

Suppose  $A_{kl} \subseteq A$  is a non-empty finite subset of parameters. Let us define an  $A_{kl}$  -indiscernibility relation on the grid D,  $R_D^{A_{kl}} = \{(d_{xy}, d_{mn}) \in D \times D | \forall a_j \in A_{kl}, f(d_{xy}, a_j) = f(d_{mn}, a_j)\}$ . Using this relation, we can describe homogeneous spatial areas, which are uniform concerning the values of the parameters belonging to the subset  $A_{kl}$ , represented by the approximating set of cells, and denoted by h. All cells that belong to the spatial area h are  $A_{kl}$ -indiscernible.

Each spatial area cannot overlap or cover one another, but they can be adjacent or adjoin to one another. They have such features as continuity and connectivity (spatial concentration of the underlying cells).

Let H be a set of spatial areas,  $H = \{h_1, ..., h_k\}$ . To represent a plurality of spatial areas that have not the property of the continuity, but describe a set of separate areas spatially distributed on the

set D, we can also use a certain  $A_{lm}$ -indiscernibility relation defined over H,  $R_H^{A_{lm}} = \left\{ \forall h_l, h_q \in H, \ \forall d_m, d_n \in D, \ \exists d_m \in h_l, d_n \in h_q \ \middle| \ \forall a_k \in A_{lm}, f\left(d_m, a_k\right) = f\left(d_n, a_k\right) \right\}$ . Obviously, all areas belonging to  $R_H^{A_l}$  are  $A_{lm}$ -indiscernible.

## 5.1. Energy transfer

The accumulation of energy inside the cell cannot continue indefinitely. As soon as the amount of energy reaches a certain predetermined level, it "overflows", forming an impulse, which should be distributed between the channels under the matrix  $\Omega$  of coefficients and transmitted to other cells.

Energy transfer is seen both as a basis of causal relationships between events and as a means of organizing them into cascading structures.

Obviously, energy can be of different types, each of which corresponds to the class of the observed event. The class of observed events, in turn, corresponds to the class of state that changed and raised the event. Usually, specific events can be driven only by the appropriate type of energy.

When the event of a certain class occurs, the impulse generator must eject some quantum of energy of a corresponding class with an appropriate amount. In general, the event model should consider the possibility that a quantum of energy contains interrelated portions of energy of different classes, which do not just correspond to the event class. The key role in understanding the dynamics of the ongoing processes is played by the dynamic parameters of the generated impulses. Depending on the impulse duration, its amplitude, the pressure it exerts on the input of the event, a different picture of the reaction of one event to the other event that has occurred can be observed.

Thus, the proposed model can transfer different types of energy by short bursts (as impulses) or by long potential inputs. The event can be only triggered if a sufficient amount of appropriate energy is received, which should be determined by the integral of the received energy over time.

Consider the portions of energy transfer. Let  $\varepsilon_k$  be an energy portion of class k that corresponds to  $w_k$ . Suppose energy portion  $\varepsilon_k$  can be described as

$$\varepsilon_k = (k, \tau, q_1, q_2), \tag{4}$$

where k is an energy class,  $\tau$  is a duration,  $q_1$  is an initial amplitude, and  $q_2$  is a final amplitude (Fig. 5).

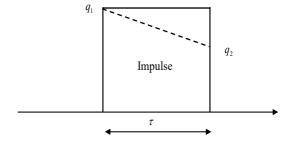


Figure 5: Dynamic parameters for energy transfer impulse

The energy quantum can be represented by a tuple of energy portions:

$$E = \langle \varepsilon_1, ... \varepsilon_m \rangle.$$
 (5)

The released energy can directly impact events sensitive to this type of energy through certain connections between them. The proposed model allows us to express not only the direct but also indirect effects of events through the energy transfer.

#### 5.2. Connectors

Let  $\varphi_l$  be a connector that connects events  $\varphi_j$  and  $\varphi_k$  transferring energy portions  $\varepsilon_{j1},...\varepsilon_{jm}$  from the first one to the second. Such transfer for each energy portion can be labeled by certain confidence represented by a gray number  $\mu_l^{\pm}$  and a certain time  $t_l^{\pm}$ , which is also grayed. The connector  $\varphi_l$  can also have a sensitivity point  $\sigma_l$ , which controls the transfer process, if necessary. Suppose the impact of a certain energy portion  $\varepsilon_l$  on the connector allows it to block, break, or amplify the energy transfer from  $\varphi_l$  to  $\varphi_k$  (depending on the type and amount of  $\varepsilon_l$ ).

Thus, the connector  $\varphi_i$  can be defined as a tuple

$$\varphi_{l} = \left\langle \left\{ \varepsilon_{i1}, \dots \varepsilon_{im} \right\}, \psi_{i}, \psi_{k}, \mu_{l}^{\pm}, t_{l}^{\pm}, \sigma_{l} : \left\{ \varepsilon_{i1}, \dots \varepsilon_{iw} \right\} \right\rangle, \tag{6}$$

where  $\varphi_j$  and  $\varphi_k$  are events,  $\{\varepsilon_{j1},...\varepsilon_{jm}\}$  is a subset of energy portions permitted to transfer through  $\varphi_l$ ,  $\mu_l^{\pm}$  is a likelihood,  $t_l^{\pm}$  is the time to receive energy portion, and  $\sigma_l$  is an optional sensitivity point with the energy portions  $\{\varepsilon_{l1},...\varepsilon_{lw}\}$  permitted to receive.

Let  $\psi_k$  be a meta-connector that connects the event  $\varphi_j$  and the sensitivity point  $\sigma_l$  of a certain connector. It allows transferring energy portions of given classes  $\{\varepsilon_{f1},...\varepsilon_{fm}\}$  with a certain degree of acceleration  $\chi_k$  as follows:

$$\psi_{k} = \left\langle \left\{ \varepsilon_{f1}, \dots \varepsilon_{fm} \right\}, \varphi_{j}, \sigma_{l}, \chi_{k} \right\rangle. \tag{7}$$

#### 5.3. Abstract Events

Suppose an abstract event  $\eta$  consists of a set of inputs  $\{x_{k_1}, x_{k_2}, ... x_{k_m}\}$ , each of which is sensitive to a class k of energy and receive only the correspondent energy potions  $\varepsilon_k$ , a set of outputs  $\{y_{k_1}, ... y_{k_n}\}$ , each of which ejects an energy portion  $\varepsilon_p$  of a class p, and a set of accumulators  $\Xi = \{\pi_{k_1}, ... \pi_{k_m}\}$ , where each  $\pi_{k_j}$  accumulates energy portions  $\varepsilon_j$  of a class j through the input  $x_{k_j}$ .

Each input  $x_{ki}$  is connected to the accumulator  $\pi_{ki}$  with a weight (multiplication factor)  $\zeta_{ki}$ .

Thus, the energy portion  $\varepsilon_j$  received through input  $x_{kj}$  adds the amount of energy  $\zeta_{kj} \cdot \int_0^1 q\tau$  of type j to the accumulator  $\pi_{kj}$ . If accumulated  $\pi_{kj}$  energy exceeds a threshold value  $\beta_{kj}$ , a certain quantum of energy E should be released and ejected to the connectors. The classes, amplitudes, and durations of energy portions released by energy quantum depend on a certain multiplication factor  $\gamma_{kj}$ .

Thus, the abstract event  $\eta_k$  can be represented as

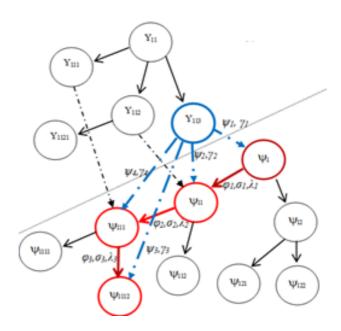
$$\eta_k = \langle X_k, Y_k, \Xi_k \rangle, \tag{8}$$

where  $X_k$  is an input part,  $X_k = \{(x_{k1}, \zeta_{k1}), ...(x_{km}, \zeta_{km})\}$ ,  $Y_k$  is an output part,  $Y_k = \{y_{k1}, ..., y_{kn}\}$ , and  $\Xi_k$  is an accumulation part,  $A_k = \{(\pi_{k1}, \beta_{k1}, \gamma_{k1}), ...(\pi_{km}, \beta_{km}, \gamma_{km})\}$ . The proposed model of the event is flexible and dynamic, since the weights  $\zeta_{kj}$ , threshold values  $\lambda_{kj}$ , and factors  $\gamma_{kj}$  can dynamically change in time.

#### 5.4. Abstract Event Network

An abstract event network can be represented as a time-ordered event structure  $G = \left\langle \left\{ \eta_k \right\}_{k=1}^n, \tau, \nu, \left\{ \varphi_l \right\}_{l=1}^w, \left\{ \psi_k \right\}_{k=1}^v \right\rangle$ , where  $\left\{ \eta_k \right\}_{k=1}^n$  is a set of abstract events,  $\tau : \eta \to T$  is a mapping that expresses sequential order of the events,  $\nu : \eta \to H$  is a mapping that expresses the spatial reference of events,  $\left\{ \varphi_l \right\}_{l=1}^w$  is a set of connectors between events, and  $\left\{ \psi_k \right\}_{k=1}^v$  is a set of metaconnectors, which connect events and corresponding sensitivity point of the connectors. This formalization allows the use of soft sets to represent different sequences of events, based on the indiscernibility relation between events by class, spatial position, or time intervals.

Fig. 6 shows the abstract event network, in which nodes represent events and the arcs represent connectors and meta-connectors.



## 6. Implementation

The proposed model has been implemented using Visual C++ based on the double indexed lists and approbated on the simulated area. The AOI is the Lower Dnieper Sands (Oleshky Sands) in the Kherson region, in Southern Ukraine. The sands are surrounded by very dense artificial coniferous forests that prevent the sands from moving during strong winds. Global warming leads to the loss of forests in this area. As a result of global warming, we can observe chains of cascading effects. Due to warming, the groundwater levels are decreased, which further increases fire danger, rapid destruction of forests in large areas, desertification of the territory, and the revival of sand movement. Due to warming, forests are also being affected by invasions of insects, and also become more prone to forest fires (Fig. 7).

The results of the conducted simulation show that the proposed model provides enough performance to real-time modeling of a wide range of natural processes from climate change to forest fires and adequate knowledge representation about cascading events taking into account the uncertainty of the observations.

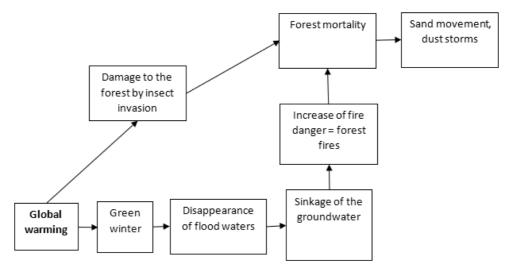


Figure 7: Cascading chains of events in Kherson Region

#### 7. Conclusion

The proposed event-based model enables knowledge representation about an observed plurality of dynamic processes of various intensities that arise unexpectedly and evolve simultaneously in a wide range of domains. It is based on events that are referenced to time intervals and spatial areas and take into account the uncertainty of the observed information about the evolving processes. Uncertainty is represented by gray numbers and soft sets. Events are connected in an abstract network, where arcs represent connectors and meta-connectors that transfer energy by impulses to model transitions of a dynamic system from state to state. The proposed abstract event network can be used to model the evolution of dynamic systems that can be expressed by

events that occurred inside the system but driven by impacts from the outside (environmental effects). Such evolution is considered in a wide range of rates, from the slowest processes associated with climate change to the most rapid processes associated with the effects of natural forces of a destructive nature. The proposed model also makes it possible to adequately represent sequential, parallel, and cascade chains of events with a trigger effect, information about which is incomplete and inaccurate. Future research will be devoted to the study of the coefficient matrices and formalization of the process of generating energy quantum.

#### References

- [1] D.Slater, Modelling, Monitoring, Manipulating and Managing? Modelling Process Flow in Complex Systems. Cambrensis: Porthcawl, UK, 2017
- [2] N.G.Leveson, Engineering a Safer World: Systems Thinking Applied to Safety, MIT Press, Cambridge, MA, 2011
- [3] T.Bjerga, T.Aven, E.Zio, Uncertainty treatment in risk analysis of complex systems: The cases of STAMP and FRAM, Reliability Engineering & System Safety, No. 156, 2016, pp.203–209. doi:10.1016/j.ress.2016.08.004
- [4] A.Adriaensen, W.Decré, L.Pintelon, Can Complexity-Thinking Methods Contribute to Improving Occupational Safety in Industry 4.0? A Review of Safety Analysis Methods and Their Concepts, Safety 5, No. 65, 2019. doi:10.3390/safety5040065
- [5] H.Slim, S.Nadeau, A proposal for a predictive performance assessment model in complex sociotechnical systems combining fuzzy logic and the Functional Resonance Analysis Method (FRAM), American Journal of Industrial and Business Management, No. 9, 2019, pp. 1345–1375. doi: 10.4236/ajibm.2019.96089
- [6] M.Zharikova, V.Sherstjuk, Threat assessment method for intelligent disaster decision support system, In: N. Shakhovska (Ed.) Advances in Intelligent Systems and Computing, Springer, Cham (2016), vol. 512, 2016, pp. 81–99. doi:10.1007/978-3-319-45991-2\_6
- [7] A.Bies, Z.Song, J.Getman, J.Ellis, J.Mott, S.Strassel, M.Palmer, T.Mitamura, M.Freedman, H.Ji, T.O'Gorman, A Comparison of Event Representations in DEFT, In: 4th Workshop on Events: Definition, Detection, Coreference, and Representation, San Diego, California, 2016, pp. 27–36. doi: 10.18653/v1/W16-1004
- [8] L.Gao, P.K.Choubey, R.Huang, Modeling Document-level Causal Structures for Event Causal Relation Identification, In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 1, 2019, pp. 1808–1817 doi: 10.18653/v1/N19-1179
- [9] Z.Jin, S.Guo, N.Chen, D.Weiskopf, D.Gotz, N.Cao, Visual Causality Analysis of Event Sequence Data, IEEE Transactions on Visualization and Computer Graphics, No. 27(2), 2021, pp. 1343–1352. doi: 10.1109/tvcg.2020.3030465
- [10] P.Gärdenfors, Events and Causal Mappings Modeled in Conceptual Spaces, Frontiers in Psychology, 11, 2020, p.630. doi: 10.3389/fpsyg.2020.00630
- [11] K.Liu, Y.Chen, J.Liu, X.Zuo, J.Zhao, Extracting Events and Their Relations from Texts: A Survey on Recent Research Progress and Challenges, AI Open, No. 1, 2020, pp. 22–39. doi: 10.1016/j.aiopen.2021.02.004
- [12] P.Baldan, A.Raffaeta, Minimisation of Event Structures, In: Proc. of 39th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, vol. 150, 2019, 30:1–30:15. doi: 10.4230/LIPIcs.FSTTCS.2019.30

- [13] C.Sinha, P.Gärdenfors, Time, space, and events in language and cognition: a comparative view, Annals of the New York Academy of Sciences, Flow of Time, 2014, pp. 1–10. doi: 10.1111/nyas.12491
- [14] X.Song, J.Ding, Y.Liu, L.Zhang, J.Wang, Modeling the Process of Event Sequence Data Generated for Working Condition Diagnosis, Mathematical Problems in Engineering, No. 693450, 2015. doi: 10.1155/2015/693450
- [15] S.Wu, S.Liu, S.Sohn, S.Moon, C.Wi, Y.Juhn, H.Liu, Modeling asynchronous event sequences with RNNs, Journal of Biomedical Informatics, No. 83, 2018, pp.167–177. doi: 10.1016/j.jbi.2018.05.016
- [16] W.Wang, F.Di Maio, P.Baraldi, E.Zio, A hybrid Monte Carlo and possibilistic approach to estimate non-suppression probability in fire probabilistic safety analysis, In: Proc. of 2nd International Conference on System Reliability and Safety (ICSRS), 2017, pp. 552–559. doi: 10.1109/ICSRS.2017.8272881
- [17] E.E.H.Doyle, D.M.Johnston, R.Smith, D.Paton, Communicating model uncertainty for natural hazards: A qualitative systematic thematic review, International Journal of Disaster Risk Reduction, No. 33(2), 2018, pp.449–476. doi: 10.1016/j.ijdrr.2018.10.023
- [18] J.Guin, Modeling Fundamentals: Understanding Uncertainty, AIR Currents 4, 2018. URL: https://www.air-worldwide.com/Publications/AIR-Currents/2018/Modeling-Fundamentals-Understanding-Uncertainty/
- [19] V.Sherstjuk, M.Zharikova, Plausible Event-Tree Networks for Knowledge Representation in Real-Time GIS-Based Decision Support Systems, CEUR Workshop Proc., No. 2105, 2018, pp.109–123
- [20] D.D.Nolte, Introduction to Modern Dynamics: Chaos, Networks, Space and Time, Oxford University Press, 2019
- [21] P.L.C.Van Geert, Dynamic Systems, Process and Development, Human Development, No. 63, 2019, pp.153–179. doi:10.1159/000503825
- [22] Z.Zhang, Statistical methods on risk management of extreme events (Statistical Methodology Commons), Doctoral Dissertations, 1002, University of Massachusetts, Amherst, MA, 2017
- [23] P.L.Guth, Military geology in war and peace: an introduction, In: Underwood, J.R., Guth, P.L. (Eds.) Geology in War and Peace, vol. XIII of Reviews in Engineering Geology, 1998. doi:10.1130/REG13
- [24] B.Wahlström, C.Rollenhagen, Safety management A multi-level control problem. Safety Science, No. 69, 2014, pp.3–17. doi: 10.1016/j.ssci.2013.06.002
- [25] N.Alon, J.H.Spencer, The Probabilistic Method, Wiley, New York, 2016
- [26] J.M.Caplan, L.W.Kennedy, J.D.Barnum, Risk Terrain Modeling for Spatial Risk Assessment, Cityscape, No. 17(1), 2015, pp.7–16
- [27] C.C.Aggarwal, Managing and Mining Uncertain Data, In: Aggarwal, C.C. (Ed.) Advances in Database Systems, Springer, Boston, vol. 35, 2010. doi:10.1007/978-0-387-09690-2
- [28] S.P.Demri, E.S.Orlowska, Incomplete Information: Structure, Inference, Complexity. Springer-Verlag, Berlin Heidelberg, 2002. doi:10.1007/978-3-662-04997-6
- [29] L.A.Zadeh, Fuzzy sets, Information and Control, No. 3(8), 1965, pp.338–353. doi:10.1016/S0019-9958(65)90241-X
- [30] S.A.H.Minoofam, J.Ahmadi, H.R.Kanan, A comparative review on nondeterministic sets for association rule mining, In: Proc. of the 4th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), Zahedan, Iran, 2015, pp. 1–5. doi:10.1109/CFIS.2015.7391691

- [31] F.Smarandache, Neutrosophic Set is a Generalization of Intuitionistic Fuzzy Set, Inconsistent Intuitionistic Fuzzy Set (Picture Fuzzy Set, Ternary Fuzzy Set), Pythagorean Fuzzy Set, q-Rung Orthopair Fuzzy Set, Spherical Fuzzy Set, and n-HyperSpherical Fuzzy Set, while Neutrosophication is a Generalization of Regret Theory, Grey System Theory, and Three-Ways Decision. Journal of New Theory, No. 29, 2019, pp.01–35
- [32] Z.Pawlak, Rough sets. International Journal of Computer & Information Sciences, No. 11(5), 1982, pp. 341–356. doi:10.1007/BF01001956
- [33] S.Liu, Y.Lin, Grey Information: Theory and Practical Applications, Springer-Verlag London, 2006. doi:10.1007/1-84628-342-6
- [34] D.Molodtsov, Soft set theory First results, Computers & Mathematics with Applications No. 37(4–5), 1999, pp.19–31. doi:10.1016/S0898-1221(99)00056-5
- [35] D.Yamaguchi, G.-D.Li, L.-C.Chen, M.Nagai, Reviewing crisp, fuzzy, grey and rough mathematical models, In: Proc. of 2007 IEEE Int. Conf. on Grey Systems and Intelligent Services, Nanjing, China, 2007, pp. 547–552. doi:10.1109/GSIS.2007.4443334
- [36] P.S.Abril, R.Plant, The patent holder's dilemma: Buy, sell, or troll?, Communications of the ACM, No. 50, 2007, pp. 36–44. doi:10.1145/1188913.1188915
- [37] S.Cohen, W.Nutt, Y.Sagic, Deciding equivalences among conjunctive aggregate queries, J.ACM, No. 54, 2007. doi:10.1145/1219092.1219093.
- [38] J.Cohen (Ed.), Special issue: Digital Libraries, volume 39, 1996
- [39] D.Kosiur, Understanding Policy-Based Networking, 2nd. ed., Wiley, New York, NY, 2001
- [40] D.Harel, First-Order Dynamic Logic, volume 68 of Lecture Notes in Computer Science, Springer-Verlag, New York, NY, 1979. doi:10.1007/3-540-09237-4
- [41] I.Editor (Ed.), The title of book one, volume 9 of The name of the series one, 1st. ed., University of Chicago Press, Chicago, 2007. doi:10.1007 3-540-09237-4
- [42] I.Editor (Ed.), The title of book two, The name of the series two, 2nd. ed., University of Chicago Press, Chicago, 2008. doi:10.1007/3-540-09237-4
- [43] A.Z.Spector, Achieving application requirements, in: S. Mullender (Ed.), Distributed Systems, 2nd. ed., ACM Press, New York, NY, 1990, pp. 19–33. doi:10.1145/90417
- [44] B.P.Douglass, D.Harel, M.B.Trakhtenbrot, Statecarts in use: structured analysis and object-orientation, in: G.Rozenberg, F.W.Vaandrager (Eds.), Lectures on Embedded Systems, volume 1494 of Lecture Notes in Computer Science, Springer-Verlag, London, 1998, pp. 368–394. doi:10.1007/3-540-65193-4 29
- [45] D.E.Knuth, The Art of Computer Programming, Vol. 1: Fundamental Algorithms (3rd. ed.), Addison Wesley Longman Publishing Co., Inc., 1997
- [46] S.Andler, Predicate path expressions, in: Proceedings of the 6th. ACM SIGACT-SIGPLAN symposium on Principles of Programming Languages, POPL '79, ACM Press, New York, NY, 1979, pp. 226–236. doi:10.1145/567752.567774
- [47] S.W.Smith, An experiment in bibliographic mark-up: Parsing metadata for xml export, in: R. N. Smythe, A. Noble (Eds.), Proceedings of the 3rd. annual workshop on Librarians and Computers, volume 3 of LAC '10, Paparazzi Press, Milan Italy, 2010, pp. 422–431. doi:99.9999/woot07-S422
- [48] M.V.Gundy, D.Balzarotti, G.Vigna, Catch me, if you can: Evading network signatures with web-based polymorphic worms, in: Proceedings of the first USENIX workshop on Offensive Technologies, WOOT '07, USENIX Association, Berkley, CA, 2007
- [49] D.Harel, LOGICS of Programs: AXIOMATICS and DESCRIPTIVE POWER, MIT Research Lab Technical Report TR-200, Massachusetts Institute of Technology, Cambridge, MA, 1978

- [50] K.L Clarkson, Algorithms for Closest-Point Problems (Computational Geometry), Ph.D. thesis, Stanford University, Palo Alto, CA, UMI Order Number: AAT 8506171, 1985.
- [51] D.A.Anisi, Optimal Motion Control of a Ground Vehicle, Master's thesis, Royal Institute of Technology (KTH), Stockholm, Sweden, 2003
- [52] H.Thornburg, Introduction to bayesian statistics, 2001. URL: http://ccrma.stanford.edu/jos/bayes/bayes.html
- [53] R.Ablamowicz, B.Fauser, Clifford: a maple 11 package for clifford algebra computations, version 11, 2007. URL: http://math.tntech.edu/rafal/cli11/index.html
- [54] Poker-Edge, Com, Stats and analysis, 2006. URL: http://www.poker-edge.com/stats.php
- [55] B.Obama, A more perfect union, Video, 2008. URL: http://video.google.com/videoplay? docid=6528042696351994555
- [56] D.Novak, Solder man, in: ACM SIGGRAPH 2003 Video Review on Animation theater Program: Part I - Vol. 145 (July 27–27, 2003), ACM Press, New York, NY, 2003, p. 4. URL: http://video.google.com/videoplay?docid=6528042696351994555. doi:99.9999/woot07-S422.
- [57] N.Lee, Interview with bill kinder: January 13, 2005, Comput. Entertain, No. 3, 2005. doi:10.1145/1057270.1057278.
- [58] J.Scientist, The fountain of youth, 2009. Patent No. 12345, Filed July 1st., 2008, Issued Aug. 9th., 2009
- [59] B.Rous, The enabling of digital libraries, Digital Libraries, No. 12, 2008
- [60] M.Saeedi, M.S.Zamani, M.Sedighi, A library-based synthesis methodology for reversible logic, Microelectron. J., No. 41, 2010, pp. 185–194
- [61] M.Saeedi, M.S.Zamani, M.Sedighi, Z.Sasanian, Synthesis of reversible circuit using cycle-based approach, J. Emerg. Technol. Comput. Syst., No. 6, 2010
- [62] M.Kirschmer, J.Voight, Algorithmic enumeration of ideal classes for quaternion orders, SIAM J. Comput, No. 39, 2010, pp. 1714–1747. URL: http://dx.doi.org/10.1137/080734467. doi:10.1137/080734467
- [63] L.Hörmander, The analysis of linear partial differential operators. IV, volume 275 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, Germany, Fourier integral operators, 1985
- [64] L. H'ormander, The analysis of linear partial differential operators. III, volume 275 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, Germany, Pseudodierential operators, 1985
- [65] IEEE, Ieee tcsc executive committee, in: Proceedings of the IEEE International Conference on Web Services, ICWS '04, IEEE Computer Society, Washington, DC, USA, 2004, pp. 21–22. doi:10.1109/ICWS.2004.64
- [66] TUG, Institutional members of the TEX users group, 2017. URL: http://www.tug.org/instmem.html
- [67] R Core Team, A language and environment for statistical computing, 2019. URL: https://www.R-project.org/.
- [68] S.Anzaroot, A.McCallum, UMass citation field extraction dataset, 2013. URL: http://www.iesl.cs.umass.edu/data/data-umasscitationfield