Synthesis of text models with information streams

Yuriy Korostil¹, Olga Korostil²
1 Maritime University of Szczecin, Department of Mathematics
70-500 Szczecin, ul. Wały Chrobrego 1–2, e-mail: j.korostil@am.szczecin.pl
2 Ukrainska Akademia Drukarnstwa, Lwów

Key words: models, semantics, synthesis, texts, grammar, objects, functioning, output

Abstract
The methods of synthesis of text models with information streams are researched. In the process of functioning of system of text models, which model systems of social objects the following are implemented: analysis of models of monitoring of information means, used by social objects, control of adequacy of models to objects and objects management. The synthesis of text models with information streams is implemented basing on use of semantic parameters. In process of text forms synthesis of models description and information streams appears necessity to output new text fragments, reflecting synthesis result. Such output processes are procedures based on use of logical schemas and grammar rules.

Introduction

Process of functioning of text models system (STM) is not only in implementation of actions, connected to management of social objects (SO), being modeled with help of text models (TM). This process supposes implementation of following functional possibilities: analysis of current state of TM, monitoring of STM, management SO, basing on use of information streams (IP) and TM, control of adequacy of current state of TM to corresponding objects SO. Monitoring of STM is necessary to determine moments of activation of actions, connected to detection of factors, influencing current state of TM, which are not caused by managing actions initiated against TM. Such factors include: procedures, connected with identifying models TM with corresponding objects SO; procedures connected to analysis of TM; procedures connected to modification of TM, which is not caused by managing actions IP etc. Because imagination about management of TM is quite wide, let us create definition of such imagination.

Definition 1. As management action on TM we mean actions which are determined by following conditions: managing action to TM for change of corresponding state of SO; managing action is implemented by external relative to TM, factors by forming appropriate IP; any managing action is caused by target, description of which is included into IP.

One of main tasks which is solved with the help of monitoring system is a task of ensuring adequacy between TM and SO. Solving that task is based on following methods: forming short and long term reverse connections between system of social objects (SSO) and STM; stand-alone analysis of TM; extending interpretations, which are related to SO. Reverse connections between SO and TM are implemented basing on following components: obligatory data about SO, which are formed from data of various structures, orientated on work with SSO; data from electronic mass media; data from other sources, which appear randomly. Monitoring system solves following tasks: analysis of data about SO and forming of image of SO, and its current state SO relative to TM; forming process of monitoring of system STM and pseudo monitoring SSO, which is in definition of process parameters, for example, value of period of one cycle of monitoring of mass media, depth of monitoring, forming additional processes if there is a necessity in their initiation; implementation of processes of modification of TM, if last appeared to be necessary because of the results of adequacy check between SO and TM.

Monitoring system (SM) makes during its function only analysis of data about SO. Basing on that
analysis a new model $TM_i$ is formed, if it was detected that new $SO_i$ object appeared, or modification of $TM_i$ is conducted, if changes of parameters in corresponding $SO_i$ were detected. Feedback which supplies data to $STM_i$ about $SO_i$, which is modeled in $STM_i$, is generally passive because capacity of activation of data transfer from $SO$ to $STM$ is quite limited. Due to existence of electronic communication means such activation can be implemented but not in determined form. This means that actions implemented by $STM_i$ lead to corresponding reaction from $SO$ through some period of time. $STM_i$ system from one side is a system of images $SO_i$ in form of $TM_i$, and from other side $STM_i$ is a system helping to make level of management actions on separate $SO_i$ objects, or on whole system $SSO_i$. Reliability of results of such modeling is determined by level of distortion or level of inadequacy between $SO_i$ and $TM_i$.

Components used in process of synthesis and synthesis of text models with information streams

As far as system $STM_i$ describes $SSO_i$, then besides methods which describe separate $SO_i$ as $TM_i$, necessary methods which describe relations between $TM_i$ and $TM_j$ in $STM_i$, which correlate with relations between $SO_i$ and $SO_j$ in $SSO_i$. Such components should not necessarily be objects separated from $TM_i$ in $STM_i$. Connections between $TM_i$ and $TM_j$ can be implemented basing on level of similarity between separate $TM_i$ and $TM_j$, or, basing on similarity of $SO_i$ and $SO_j$ which is obvious for $SSO_i$. Such relation is implemented basing on following: level of structural similarity between separate $TM_i$; level of functional dependence if $TM_i$ and relative $SO_i$ are implementing predefined processes in $SSO_i$; level of dynamic similarity which can be in functional similarity of semantic; level of activation or other parametric similarity [1, 2].

Let us review method of description of structural connections which are implemented as separate text elements $STM_i$, which will be marked $TZ_i$ [3, 4]. Keeping in mind imagination about text methods of description let us accept that $TZ_i$ is some text form which dislike model $TM_i$ identify not object $SO_i$ but describes conditions of activation of connection between $TM_i$ and $TM_j$. Physical implementation of such connection is in transmission of data from one model $TM_i$ to other $TM_j$. Function of component $TZ_i$ is implemented by following steps: $TM_i$ and $TM_j$ are determined, which can implement relation in form of exchange of information; basing on analysis of $TM_i$ data for information package $IP_i^{TM}$ are selected in $TM_i$; corresponding $IP_j^{TM}$ is transmitted to $TM_j$ as information extension in framework of $TM_j$ is made corresponding modification of $TM_j$.

To analyze more deep processes of functioning, implemented by $TZ_i$ which can be written down as: $TM_i^* = TZ_i(TM_i, TM_j)$, it is necessary to stop on structure of $TM_i$ in projection on subject area, which is described by $STM_i$ and separately by $TM_j$. In that case structure of $TM_i$ will be characterized by following aspects: own structure aspects; structure aspects connected to subject area; general aspects of structure $TM_i$.

Existence of structure in $TM_i$ allows to form some rules of conduction of synthesis of $TM_i$ and $IP_i$, when $IP_i$ extends $TM_i$ in form of usual concatenation $J(TM_i)$. Obviously $IP_i$ must be isomorphic to $TM_i$ not only at the level of language, used to describe $TM_i$, but also at level of structure and rules of its forming [5, 6]. In that case it can talk about following rules of analysis which are used on separate step of functioning of $STM_i$, which is defined or activated by management action of $IP_i$.

PR1: Determination of difference between separate fragments of interacting text descriptions can be formally described by following correlation:

$$[S_i^p(j(tm_i)) = S_i^p(j(ip_i)) \rightarrow j(tm_i) - j(ip_i) = R_i^p(tm_i, ip_i)]$$

(1)

where: $S_i^p(j(tm_i))$ and $S_i^p(j(ip_i))$ – structural characteristics of fragment $tm_i$ and $ip_i$, correspondingly, $R_i^p(tm_i, ip_i)$ – level of structural correspondence of two interacting objects in $STM_i$, which are $TM_i$ and $IP_i$.

PR2: Synthesis of two fragments at level of their phrases is described by following correlation:

$$[R_i^p(tm_i, ip_i) \leq \alpha(TM_i) \rightarrow [tm_i(\varphi_1 \cdots \varphi_m) \& \varphi_p, \varphi_1 \cdots \varphi_p) = tm_i^*(\varphi_1 \cdots \varphi_m)] \lor$$

$$[R_i^p(tm_i, ip_i) > \alpha(TM_i) \rightarrow [tm_i^* = (tm_i \ast ip_i)]]$$

(2)

where $\alpha(TM_i)$ – threshold of allowed semantic difference between fragments $tm_i$ and $ip_i$, $\varphi_p$ – phrase, which comes out of $tm_i$, $\varphi_p^*$ – phrase which comes out of $ip_i$, $\varphi_j$ – phrase which comes out of $tm_i$, $tm_i^*$ – fragment which is synthesized at level of phrases from $tm_i$ and $ip_i$, $\ast$ – sign of concatenation of two text fragments $tm_i$ and $ip_i$ at level of phrases.

RP3: Rule of reduction of semantically exceeding phrases from text fragment $tm_i^*$, which is formally written down as following correlation:
Synthesis of text models with information streams

\[ \exists \left[ \phi^m, \phi^n \right] \in \text{tm}^* \left| \sum_{i} \left( S^c \left( \phi^m \right) - S^c \left( \phi^n \right) \right) \leq \alpha(\text{tm}) \right] \rightarrow \left[ \left( \frac{\text{tm}^*}{\phi^m} \right) \rightarrow \left( \phi^n \not\in \text{tm}^* \right) \right] \quad (3) \]

where \( S^c(\phi) \) – value of semantic significance of phrase \( \phi \), \( \alpha(\text{tm}) \) – boundary allowed value with which semantic difference between two phrases is allowed or not allowed.

In process of synthesis \( TM_i \) from \( IP_i \) can appear necessity of output of new phrase. Then with the aim of forming phrase \( \hat{\phi}(TM_i) \), which semantically will be equivalent to phrases \( \phi(\text{tm}_j) \) and \( \phi(ip) \).

Definition 2. Semantically equivalent phrases \( \phi \) and \( \phi \) are phrases, for which semantic conformity is greater then threshold value \( \Delta \sigma^c \).

In case of use of imagination about semantic conformity, it is necessary to take into account following peculiarity of text representation of information which is in fact that semantic content is influenced not only by level of average semantic significance of words in two different fragments, but also a place of some words in corresponding phrases. Some meaning of semantic controversy is accepted as not allowed meaning of its value between words which are part of the same phrase in correlation:

\[ \sigma^S(\phi, \phi) = \left| \sigma^S(\phi) - \sigma^S(\phi) \right| \quad (4) \]

Level of semantic conformity can be divided into two types, one of which is a general conformity defined basing on difference \( \sigma'(\phi, \phi) \) relatively to preset threshold of allowed significance of value \( \sigma'(\phi, \phi) \). It does not depend on method of placement of words in \( \phi \) and \( \phi \), as semantic conformity significantly exceeds value of controversy which is set by word replacement. In framework of \( \sigma'(\phi, \phi) \) appears necessity to take into account influence of words order in \( \phi \) and \( \phi \), to define level of conformity. Let us review method of definition of \( \Delta \sigma^c \), which divides \( \sigma'(\phi, \phi) \) into \( \sigma^c \) and \( \sigma' \). Such division is level of conformity of curve which interpolates change of value of meaning \( \sigma'(\phi) \) in framework of one phrase. In that case corresponding line is built on plane in which \( Ox \) axis reflects words in order of their placement in phrase \( \phi \). Each point on \( Ox \) which is equal \( x \) reflects \( x_i \in \phi \), and \( \Delta x = x_{i+1} - x_i \) means following word \( x_{i+1} \). On axis \( Oy \) is set value \( \sigma^S(x_i, x_{i+1}) \), which is integer, as \( \sigma'(x_i, x_{i+1}) = \left| \sigma(x_{i+1}) - \sigma(x_i) \right| \) – where corresponding values can be integers if it is accepted to define \( \sigma'(x) \) by number of words, used to describe interpretation \( x \) in semantic vocabulary \( S_C \). In case of other approaches to definition of value of \( \sigma'(x) \), for example approach basing on definition of frequency of use of word \( x \) in texts describing \( TM \) and \( IP \), which interact with them during definite period of time then value \( \sigma'(x) \) can be fractional or rational. Semantic controversy between phrases \( \phi \) and \( \phi \) is defined according to following correlation:

\[ \sigma^S(\phi, \phi) = \left| \sum_{i=1}^{m} \sigma^S(\phi) - \sum_{n=1}^{n} \sigma^S(\phi) \right| \quad (5) \]

During definition of semantic conformity \( \sigma'(\phi, \phi) \) it is necessary to take into account semantic differences between \( \phi \) and \( \phi \), which are conditioned by replacing words for extension or modification of semantics during building of phrases in which semantic controversy is prohibited. In that case, \( \sigma'(\phi, \phi) \) beside general value of \( \sigma'(\phi, \phi) \), which is medium value, must account the above change of semantics which will be called objective change of semantic parameters value. Each segment presenting relation \( \sigma'(x_i, x_{i+1}) \) in phrase \( \phi \), must have same angle as connection described by \( \sigma'(x_i, x_{i+1}) \) for phrase \( \phi \). Formula for definition \( \sigma'(\phi, \phi) \) will be:

\[ \sigma'(\phi, \phi) = \sigma^S(\phi, \phi) + \left| \sum_{i=1}^{m-1} \alpha_i - \sum_{j=1}^{n-1} \alpha_j \right| \quad (6) \]

Due to above formula it gets possible basing on a priory data to determine value of possible threshold \( \sigma^S \), which divides \( \sigma^S \) from \( \sigma' \). Due to use of imagination on interpolation curves, value of equivalence of two phrases can be determined not only as summary or middle deviation of values \( \alpha \) with \( \Gamma (\phi) \) and \( \Gamma ^c (\phi) \), but also as local parameters, localization of which is implemented basing on binding coefficients \( \alpha_i \) and \( \alpha_j \) to order number of words in phrases \( \phi \), \( \phi \). In that case it can determine maximum adequate value \( \sigma^S \). Formula to determine such value \( \sigma^S \) will be:

\[ \sigma^S = \sigma^S + \sum_{i=1, j=1}^{m,n} \alpha_i(x_{i+1}, x_j) - \alpha_j(x_{i+1}, x_j) \]

(7)

In that case it can get dependency \( \sigma^S = f(x_i, x_j) \), where \( x_i, x_j \) are coordinates in \( \phi, \phi \).

Method of output of phrases in text models

Changes taking place in \( SO \) must also take place in \( TM \). Speaking about functioning of \( TM \), it means functioning of \( \{TM, SO\} \) system. To activate processes of modification or any other changes
in \( TM_i \), excluding those changes which take place under influence of \( IP_i \), can be defined as follows: changes in objects \( SO_i \); optimization processes in \( TM_i \); results of current analysis of system \( \{ TM_i, SO_i \} \). Peculiarity of objects \( SO_i \) is in fact that they do not have direct influence on \( TM_i \), as there is no direct links between \( SO_i \) and \( TM_i \). Such relations are information only. So, reaction of \( TM_i \) on changes in \( SO_i \) can be quite complex and can be characterized by following peculiarities: reaction of \( TM_i \) on changes in \( SO_i \) can have different value of delay, as it is accepted that \( SO_i \) and \( TM_i \) are informational stand-alone; such reaction can have different level of adequacy relatively to real changes in \( SO_i \); reaction of \( TM_i \) on changes in \( SO_i \) can have different level of distortion of one or other changes or states to which comes \( SO_i \) as a result of initiation of internal modifications. Specific characteristics of \( TM_i \) and \( STM_i \), in general is a level of informational masking of \( TM_i \) relatively to \( SO_i \), level of formality of \( TM_i \) relatively to \( SO_i \), or level of completeness of reflection of \( SO_i \).

Basing on peculiarities of systems like \( SSO \), fact of existence or use of systems \( STM_i \) must be informational masked [7, 8]. Informational masking in that case means following. Fact of possibility of creation and use of systems \( STM_i \) can be known in general. But various methods of implementation of specific \( STM_i \) must not be available or known to \( SSO_i \). This parameter is a key one due to following: if it is not used then there could be possibility of direct control of objects \( SO_i \), and directly own whole information on \( SO_i \), and absence of that parameter causes necessity of full control of the whole system \( SSO_i \) that from the point of view of natural conditions of function of \( SSO_i \) is a negative factor. Formalization of description of \( TM_i \) is fact that elements which are supposed to be formally described are joined into appropriate classes and then the whole class of objects is marked in some formal way, mostly by various symbols [9, 10].

One of basic functions of system \( STM_i \) is detection of various critical situations in \( SO_i \) and detection of processes which lead \( SO_i \) to such situations. After detection of critical situations, system \( STM_i \), at least must activate processes of informing external members of modeling about that. Besides, as functions which can be implemented in \( STM_i \), there can be function of counteraction critical situations. As modeling means \( TM_i \) do not have direct influence on \( SO_i \) then appropriate functions are in forming \( IP_i \), which contain information capable to initiate counteraction of critical situations escalation in \( SO_i \). Such \( IP_i \) can be transmitted to external means orientated on execution of such influence directly guiding formed stream to corresponding \( SO_i \) to initiate elimination of critical situations in \( SO_i \). Obviously \( STM_i \) forms streams that are related to \( SO_i \) in general but not streams which could be guided to separate components of \( SO_i \), as analysis in \( TM_i \) is made only regarding \( SO_i \). In mentioned cases need of use of methods of output of new phrases \( \varphi_i \) appears, as activation of process of functioning \( SO_i \) and respectively, \( TM_i \) can require new text descriptions. Let us review some approaches to build output of phrase \( \varphi_i \) from some totality of phrases \( \{ \varphi_1, \ldots, \varphi_n \} \). In most cases such totality forms one sentence \( \psi_i \) or one paragraph \( \pi_i \).

Procedure of output of text fragments like in most of cases, especially logical, represents itself as sequence of elementary transformations [11, 12]. Despite logical schemas, during output of phrases in text environments there are following peculiarities.

1. On each output step before its realization following types of analysis are made: conducted analysis of semantic parameters of two elements between which transition is implemented, which is interpreted as one step, let us formally describe it as:

\[
[\varphi_1, \ldots, \varphi_n] \rightarrow [\varphi_1^*, \ldots, \varphi_n^*] \tag{8}
\]

where \( \varphi_i \) – random phrase of sending output step, \( \varphi_i^* \) – phrase, which arise in environment as a result of one output step; analysis of grammar correlations between words in new construction of phrase \( \varphi_i^* \) is made, which is formally written down as:

\[
\Gamma_i(x_1, \ldots, x_{ik}) \Rightarrow \gamma_i(x_1^*, \ldots, x_{ik}^*) \tag{9}
\]

where \( \Gamma_i \) – grammar rules, used in natural language of text models, \( \gamma_i \) – separate grammar rule \( \gamma_i \in \Gamma_i \), which is schema of use of separate types of words during constructions of phrase \( \varphi_i^* \); check of built phrase \( \varphi_i^* \) is made if it comply to requirements of normalization, which formally can be written down as follows: \( \lambda(\varphi_i^*) \rightarrow N(\varphi_i^*) \), where \( \lambda \in \Lambda \) – system of rules of normalization of structure of phrase or paragraph, \( N(\varphi_i^*) \) – normalized form of description of phrase \( \varphi_i^* \).

2. Implementation of output step is in use of one of operations, to which belong: elimination of words from phrase; replacement of one or couple of words with another word or group of words; adding word to phrase which is supposed to be transformed during output; changing places of words in phrase.

3. After execution of section 2 all checks described in section 1 are implemented against new phrase, and they are conducted at all levels of trans-
formed elements hierarchy, for example: \((\text{level } \psi^*_a) \rightarrow (\text{level } \psi^*_b)\), where \(\psi^*_b\) – sentence, containing created phrase.

4. Sections 1, 2, 3 are repeated until output process is complete.

Basing on analysis conducted according to sections 1, 2, 3 are formed some conditions which are considered during implementation of step of evaluation of such analysis on definite criteria. In most simple case for decision making such evaluations could be fixed binary bounds for all values of parameters or characteristics being analyzed. In cases of analysis of text forms of information presentation, making decision basing on such results is sufficient.

Let us review analysis of semantic parameters of text fragments at level of phrases \(\psi\). First let us write down general analysis procedure \(\{\phi_i, \psi_j\}\):

\[
\left(\phi_i \& \phi_j\right) \rightarrow |\sigma^S(\phi_i, \phi_j) \leq \delta \sigma^S(\phi_i \& \phi_j)| \vee \\
\left[\sigma^S(\phi_i, \phi_j) < \delta \sigma^S(\phi_i \& \phi_j)\right] \rightarrow \exists(\phi_i \& \phi_j) \rightarrow \phi_{i+1} \& \phi_{j+1} \tag{9}
\]

Semantic analysis mostly begins from definition of value of semantic controversy \(\sigma^S(\phi_i, \phi_j)\). If it is less then \(\delta \sigma^S(\phi_i, \phi_j)\), then takes place transition to next steps of analysis. If \(\sigma^S(\phi_i, \phi_j)\) is greater then \(\delta \sigma^S(\phi_i, \phi_j)\), then analysis is conducted \(\sigma^S(\psi_i, \psi_j)\), which includes appropriate phrases. If \(\sigma^S(\psi_i, \psi_j)\) is less then threshold \(\delta \sigma^S(\psi_i, \psi_j)\), then next phrase \(\phi_{i+1}\) is selected instead of \(\phi_i\). In that case we accept that \(\phi_i\) is element from \(TM_i\), and \(\phi_j\) is element from \(\psi_j\), where \(\psi_j\) is sentence from \(IP\), of some source \(IP\), (\(DIP\)), or element \(IP\), from feedback channel, which in difference from \(DIP\), we will call \(KIP\). Corresponding transition from \(\phi_i\) to \(\psi_j\) is continued until level of paragraph \(\pi_i\), which can be written down as correlation: \(\phi_i \rightarrow \psi_j \rightarrow \pi_i\), if it would appear that takes place:

\[
|\sigma^S(\pi_i, \pi_j) > \delta \sigma^S(\pi_i)| \rightarrow \\
\left[\left(KIP \rightarrow KR(SO)\right) \lor \left[KIP \rightarrow KR(DIP)\right]\right] \tag{10}
\]

Value \(\sigma^S(\pi_i, \pi_j)\) can be formed for subject area \(W_i\), or for each separate interaction of \(TM\) with \(IP\), or \(KLI\), where \(KLI\) – channel stream of information and can depend on level of necessary sensitivity \(TM\), to changes which take place under influence of \(IP\), which like \(KLI\) we will call \(DI\).

Phrase \(\phi_i\) received on previous stage is analyzed for compliance to requirements of normalization. If rule of absorption of phrase uses operations of adding words and replacement of words, then rule of normalization uses operations of elimination and replacement of words. In process of analysis of normalization requirements for words or word pairs their semantic values are determined. If \(\sigma^S(x^*_i) - \sigma^S(x^*_{i+1}) \leq \Delta \sigma\), then \(x^*_i\) and \(x^*_{i+1}\) are checked for excessiveness. Such check is in calculation of same words in \(j(x^*_i)\) and \(j(x^*_{i+1})\), which are located in semantic vocabulary \(S_C\). If number of different words with \(j(x^*_j)\) and \(j(x^*_{j+1})\) is less than some threshold \(e[j(x^*_j), j(x^*_{j+1})]\), then words \(x^*_j\) and \(x^*_{j+1}\) in framework of system of normalization \(\Lambda\) are accepted as synonyms and elimination of one of the words \(x^*_j\) or \(x^*_{j+1}\), which belong to \(TM\), is made. This circumstance is important because it has characteristics of renewing of word reserve which is used in \(TM\). According to \(\lambda(x^*_j, x^*_{j+1})\) it can appear that in \(S_C\) exists \(x^*_{j+1}\) so that exists correlation:

\[
\left\{\left[|e[j(x^*_j) - j(x^*_{j+1})]| \leq \delta \lambda_j\right) \& \\
\left(\exists x^*_{j+1} \in S_C \left[|j(x^*_{j+1}) - j(x^*_j)| < \delta^2 \lambda_j\right) \& \\
\left(j(x^*_{j+1}) - j(x^*_j) < \delta^2 \lambda_j\right) \& \left(\delta^2 \lambda_j > \delta^2 \lambda_j\right) \& \\
\delta \lambda_j > \delta^2 \lambda_j \rightarrow \left(x^*_j, x^*_{j+1}\right) \rightarrow x^*_{j+1}\right\} \tag{11}
\]

Then two words \(x^*_j\) and \(x^*_{j+1}\) are eliminated, and instead of them is used word \(x^*_{j+1}\).

**Conclusions**

Method of synthesis of text models with text information streams which are oriented on performing managing actions on social objects which are described by text models is developed. Analysis of components of process of text models functioning to which belong: analysis of models; monitoring of mass media, used by social objects; analysis of implementation of processes of model modification etc. is made. The developed methods of synthesis of text models with text images of information streams are based on use of structural characteristics of models and streams and also are based on use of semantic parameters of text images.

It is shown that during synthesis process arises necessity to implement processes of output of new text fragments which own interpretation in subject area of practical tasks of social objects management. Procedure of output of text fragments based on use of logical schemas is developed, interpretation of which does not contradict data, presented in subject area of task and is based on use of semantic parameters which characterize text forms of model presentation.
Main aim of use of text models is description of social objects which are hard to be described in formal way at necessary level of details which is a requirement to effective management of such objects. Results mentioned in work illustrate possible approach to tasks of automation of processes of synthesis and managing social objects.

References