Chemical Technology, Control and Management

Volume 2018
Issue 3 *SPECIAL ISSUE 4-5*

Article 24

10-19-2018

Implementation of the fuzzitic semantic hypernet based on the graphic models to provide the reliability of information in the systems of electronic document circulation.

A.R Akhatov
Department of Information Technology, Samarkand State University, Uzbekistan, Address: University blvd., 140104, Samarkand city, Republic of Uzbekistan, akmalar@rambler.ru

Recommended Citation
Akhatov, A.R (2018) "Implementation of the fuzzitic semantic hypernet based on the graphic models to provide the reliability of information in the systems of electronic document circulation.," *Chemical Technology, Control and Management*: Vol. 2018 : Iss. 3 , Article 24.
DOI: https://doi.org/10.34920/2018.4-5.108-113
Available at: https://uzjournals.edu.uz/ijctcm/vol2018/iss3/24

Implementation of the fuzzitic semantic hypernet based on the graphic models to provide the reliability of information in the systems of electronic document circulation.

Cover Page Footnote

Tashkent State Technical University, SSC «UZSTROYMATERIALY», SSC «UZKIMYOSANOAT», JV «SOVPLASTITAL», Agency on Intellectual Property of the Republic of Uzbekistan

ISSN 1815-4840
CHEMICAL TECHNOLOGY. CONTROL AND MANAGEMENT
2018, Special issue №4-5 (82-83) pp.108-113. https://doi.org/10.34920/2018.4-5.108-113
International scientific and technical journal
journal homepage: https://uzjournals.edu.uz/ijctcm
Since 2005

IMPLEMENTATION OF THE FUZZITIC SEMANTIC HYPERNET BASED ON THE GRAPHIC MODELS TO PROVIDE THE RELIABILITY OF INFORMATION IN THE SYSTEMS OF ELECTRONIC DOCUMENT CIRCULATION

A.R.Akhatov

Department of Information Technology, Samarkand State University, Uzbekistan
Address: University blvd., 140104, Samarkand city, Republic of Uzbekistan
E-mail: akmalar@rambler.ru

Abstract. *The problem is formulated for increasing the reliability of information of electronic documents (ED). The methods of increasing the reliability of information in electronic document circulation systems are investigated on the basis of search, recognition and classification of requisites, attributes, ED formats for the purpose of detection and correction of distortions. A semantic hypernet was designed on the basis of synthesis of models of stochastic search, fuzzy logic, methods of formation and use of knowledge bases (KB) with fuzzy rules and databases. The software complex is implemented as part of electronic document circulation systems.*

Key words: *electronic document, distortion, reliability, reference verification, image, signal characteristics, document structure, hypernet, graph model, document flow***.**

Introduction

In modern conditions of the economy based on innovative knowledge, new principles, methods, and management technologies are becoming highly demanded for successful competitive activity, allowing to optimize the functionality of providing systems based on new information technologies, mathematical and economic modeling [1].

In this connection the technologies of software services for various purposes are used to optimize the functioning of electronic document circulation systems (EDS), information resource networks (IRS), monitoring systems for production and technological indicators, etc. EDS technologies in a wide class of methods, software and hardware are presented by editors of natural language spelling, machine translation, synthesis, analysis, generation and translation of texts, performance monitoring of organizing-administrative documents. In EDS, many well-known applications of Microsoft, Lucent, Lernout & Hauspie, Unisys, Elan, etc. are also used [2].

In the present work, analysis of known technologies is conducted to select the direction of a new specific study on the problem of increasing the reliability of ED information on the basis of the fuzzy semantic hypernet (FSGN) model, as well as the results of development of a software package aimed at increasing the reliability of information in the FSGN based on search, recognition and classification, control of the authenticity of requisites, attributes, ED formats for purpose of detecting and correcting distortions.

1. Construction of semantic hypernet based on a graph model

In practice, there is a need to solve various optimization problems described by the graph model. Among them, it is important to build a fuzzy hypernet for searching, extracting and using fuzzy rules to control the reliability of information in electronic texts of computer-based workflow of enterprises using databases (DB), frequency dictionaries, grammatical structured models of natural language word forms and knowledge bases (KB) based on the use of models of fuzzy inference and neural networks [3]. In this case, the oriented fuzzy graph is constructed for searching and controlling the distorted word form with aim of correcting errors in texts.

The key point of the research is optimization of the fuzzy graph model to control the reliability of textual information in the structure of electronic documents by fuzzy criteria, in which the vertices of the graph, edges of the graph, vertices and edges of the graph are considered as variables. We consider three variants of the problem solution.

Variant 1. A fuzzy graph model of the form $\tilde{G}_1 = (\tilde{X}, F)$, in which, $\tilde{X} = \{< \mu(x_i)/x_i >\}$ is the fuzzy set of graph vertices; $i \in I = \{1,2,...,n\}$ is the set of edges of the graph; F is the value of the membership function (MF) μ for a vertex $x_i \in X$, $\mu : X \to [0,1]$ is projected.

Variant 2. The distances between the vertices of the graph are given by fuzzy numbers. At the same time, fuzzy graph model of the form $\tilde{G}_2 = (X, \tilde{F})$ is designed, in which, $X = \{x_i\}$, $i \in I = \{1,2,...,n\}$ is the set of graph vertices; , $\tilde{F} = \{< \mu_{\tilde{F}} < x_i, x_j > / < x_i, x_j >>\}$ is the fuzzy set of graph's edges; $\mu_{\tilde{F}} < x_i, x_j >$ is the value of MF $\mu_{\tilde{F}}$ for the edge $< x_i, x_j >$ and $\mu_{\tilde{F}} : X^2 \to [0,1]$.

Variant 3. A fuzzy graph model of the form $\tilde{G}_3 = (\tilde{X}, \tilde{F})$, in which, $\tilde{X} = \{< \mu(x_i)/x_i >\}$, $i \in I = \{1,2,...,n\}$ is the fuzzy set of graph's vertices, $\mu(x_i)$ is the value of the MF for the vertex $x_i \in X$, $\mu : X \to [0,1]$.; $\tilde{F} = \{< \mu_{\tilde{F}} < x_i, x_j > / < x_i, x_j >>\}$, $x_i, x_j \in X$ is the fuzzy set of graph's edges of the, $\mu_{\tilde{F}} < x_i, x_j >$ is the value of $\mu_{\tilde{F}}$ for the edge $< x_i, x_j >$ and $\mu_{\tilde{F}} : X^2 \to [0,1]$ is projected.

To determine the parameters of graph models we use the following definitions. The search path ρ is represented by a sequence of arcs $(a_1, a_2, ..., a_q)$, as its length a fuzzy number $\tilde{l}(p)$ equal to the sum of the lengths of all arcs entering into ρ, i.e.

$$\tilde{l}(p) = \sum_{(x_i, x_j) \in p} \tilde{c}_{ij}$$

where $\tilde{c}_{ij}$ is the fuzzy number of the length of the arc joining the vertices x_i and x_j.

To determine the shortest fuzzy path between the vertices x_i and $x_k \in X$ the fuzzy number $\overline{p}$ is accepting for which the following is true

$$\tilde{l}(\overline{p}_{ik}) = \min_r \tilde{l}_r(p_{ik}),$$

where p_{ik} is the path between vertices $x_i, x_k \in X$; $r = 1,2,..., s$,

s is the number of different paths between vertices $x_i, x_k \in X$ of graph $\tilde{G}$.

The path between the vertices is represented in the form of fuzzy numbers and when the MF is determinated, they are segmented.

Let two fuzzy numbers $\tilde{A}$ and $\tilde{B}$, MF $\mu_{\tilde{A}}(y)$ and $\mu_{\tilde{B}}(y)$ be given and they are represented as a set of segments corresponding to certain levels α_k, $k = 1,...,n$, i.e.

$$\mu_{\tilde{A}}(y) = \{[l_{\tilde{A}}^{\alpha_k}, r_{\tilde{A}}^{\alpha_k}]\},$$

$$\mu_{\tilde{B}}(y) = \{[l_{\tilde{B}}^{\alpha_k}, r_{\tilde{B}}^{\alpha_k}]\},$$

where $l_{\tilde{A}}^{\alpha_k}$, $r_{\tilde{A}}^{\alpha_k}$, $l_{\tilde{B}}^{\alpha_k}$, $r_{\tilde{B}}^{\alpha_k}$ respectively, the left and right boundaries of the confidence interval of the level α_k of numbers $\tilde{A}$ and $\tilde{B}$.

Addition of two fuzzy numbers is carried out as

$$\tilde{A} + \tilde{B} = \{[l_{\tilde{A}}^{\alpha_k} + l_{\tilde{B}}^{\alpha_k},\ r_{\tilde{A}}^{\alpha_k} + r_{\tilde{B}}^{\alpha_k}]\}.$$

Multiplying the fuzzy number by a factor is carried out as:

$$\tilde{A} \cdot v = \{[v \cdot l_{\tilde{A}}^{\alpha_k},\ v \cdot r_{\tilde{A}}^{\alpha_k}]\}.$$

To perform comparison of fuzzy segment numbers for each number, the ranking index is calculated. Then, the obtained ranking indexes are compared. The smaller of them will be that number whose index of ranking is less.

There are many ranking indexes for determining the minimum of two fuzzy numbers. The simplest of them is an index corresponding to the value of the maximum reliability, i.e.

$$I_1(\tilde{A}) = l_{\tilde{A}}^{\alpha=1} = r_{\tilde{A}}^{\alpha=1}.$$

Another way to identify the parameter of the graph model is to use fuzzy triangular numbers. Their peculiarity is that each fuzzy triangular number is determined by three values: the first (d_s) is less than which can not be; the second (d_l) is more than which can not be, and, finally, the third (d_m), which determines the maximum level of belonging. Let's consider operations on fuzzy numbers of a triangle.

Suppose two fuzzy numbers are given: $\tilde{A} = (d_{s1}, d_{m1}, d_{l1})$ and $\tilde{B} = (d_{s2}, d_{m2}, d_{l2})$. Their sum determined as follow

$$\tilde{A} + \tilde{B} = (d_{s1} + d_{s2}, d_{m1} + d_{m2}, d_{l1} + d_{l2}).$$

Multiplying the coefficient v by fuzzy number $\tilde{A} = (d_{s1}, d_{m1}, d_{l1})$ of a triangle are carried out by formula

$$\tilde{A} \cdot v = (v \cdot d_{s1}, v \cdot d_{m1}, v \cdot d_{l1}).$$

For comparison of fuzzy triangles, a method is used which reflects the center of gravity of the triangle, i.e.

$$C(\tilde{A}) = \frac{d_{s1} + d_{m1} + d_{l1}}{3}.$$

The method takes into account the location of the number on the real axis and the shape of MF.

Now, to determine each vertex $x_i \in X$, we will offer two fuzzy gear ratios.

In accordance with each of the above options for constructing a graph model, we obtain the following expressions for calculating fuzzy gear ratios.

For variant 1:

$$\tilde{\sigma}_0(x_i) = \sum_{x_j \in X} \tilde{d}(x_i, x_j),\ \tilde{\sigma}_t(x_i) = \sum_{x_j \in X} \tilde{d}(x_j, x_i),$$

$$(v_j = 1).$$

For variant 2:

$$\tilde{\sigma}_0(x_i) = \sum_{x_j \in X} \tilde{v}_j d(x_i, x_j),$$

$$\tilde{\sigma}_t(x_i) = \sum_{x_j \in X} \tilde{v}_j (x_j, x_i).$$

For variant 3:

$$\tilde{\sigma}_0(x_i) = \sum_{x_j \in X} \tilde{v}_j \tilde{d}(x_i, x_j),$$

$$\tilde{\sigma}_t(x_i) = \sum_{x_j \in X} \tilde{v}_j \tilde{d}(x_j, x_i).$$

In these formulas $\tilde{v}_j$ is the fuzzy weight of the vertex x_j, which characterizes the significance of the vertex $(\tilde{v}_j \in [0,1])$;

$v_j d(x_i, x_j)$ are weighted path lengths from vertex x_i to vertex x_j;

numbers $\tilde{\sigma}_0(x_i)$ and $\tilde{\sigma}_t(x_i)$ correspond to fuzzy outer and inner gear numbers of the vertex x_i.

For vertex $\bar{x}_0$ the external median of the graph $\tilde{G}$ is determined as

$$\tilde{\sigma}_0(\bar{x}_0) = \min_{x_i \in X} [\tilde{\sigma}_0(x_i)],$$

and for vertex $\bar{x}_t$ the internal median of the graph $\tilde{G}$ is determined as

$$\tilde{\sigma}_t(\bar{x}_t) = \min_{x_i \in X} [\tilde{\sigma}_t(x_i)].$$

For vertex of the graph $\bar{x}_{0,t}$ external and internal medians are determined as the sum of external and internal gear ratios

$$\tilde{\sigma}_{0,t}(\bar{x}_{0,t}) = \min_i (\tilde{\sigma}_0(x_i) + \tilde{\sigma}_t(x_i)).$$

The developed scientific and methodological foundations for constructing a fuzzy hypernet for search, placement of search images are intended for implementation in the control system of the reliability during transmission of electronic documents texts when of document circulation of enterprises and organizations is automated.

The offered method for designing a hypernet allows to optimize the process of verifying the reliability of ED information with various graph model architectures for implementation in electronic document circulation systems based on models of fuzzy conclusions, knowledge bases and databases.

2. The program complex for increasing the reliability of information and analyzing its effectiveness

In order to create a technology to increase the reliability of information, a workflow model has been developed and it is focuses on the use of UML diagrams of ED communications, which allows to reflect the links between the structural components of organizational and administrative documents (OAD) adopted as basic in building an EDS [4,5]. In addition, the offered workflow schemes contribute to the effective design of information and data models, on the basis of which methods and algorithms for searching, recognizing and controlling the reliability of ED structural elements are built. The information model of the workflow with the UML diagram is used to detail the ED structure in order to reflect the pattern of interaction and interrelations of the elements for the formation of databases and knowledge (DB and KB), the presentation of the object and functional parts of the EDS [6].

Implemented methods and algorithms for data processing are designed to solve the problems of optimizing the allocation and fixing of flows in accordance with the tasks of ED processing, the possibility of computing resources, the load of composite program modules, priorities, restrictions on the duration of processing and execution of the OAD.

Figure 1 shows the scheme of the functioning of the software on the basis of the integration of information processing by ED.

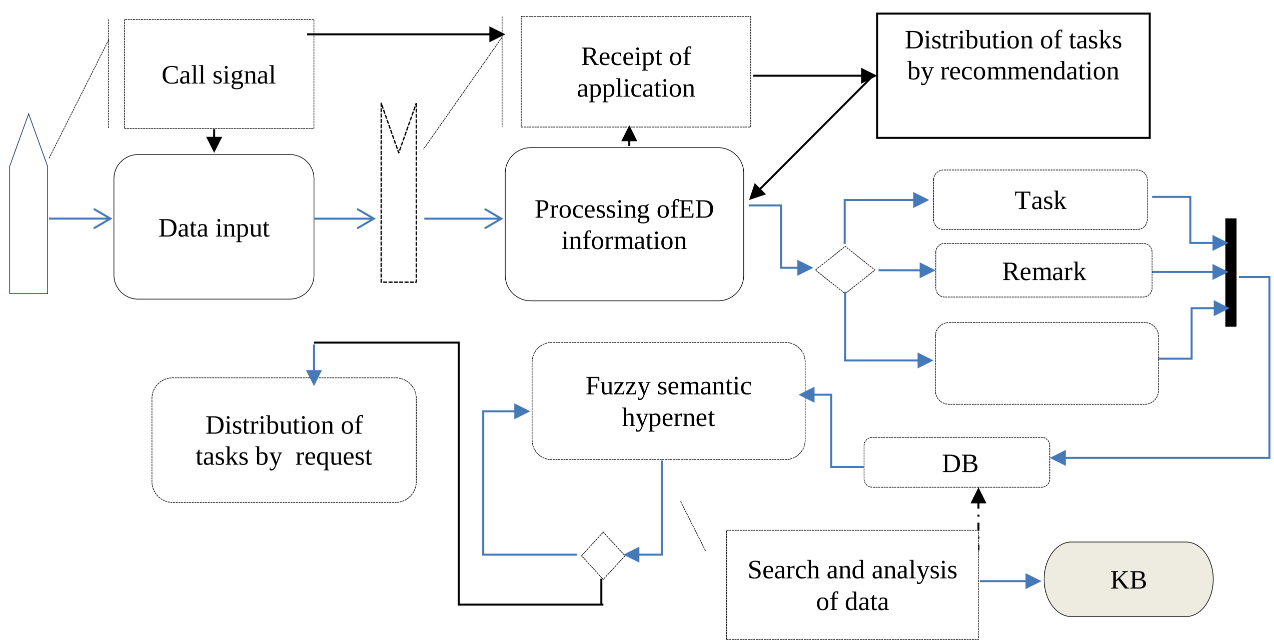

Fig. 1. ***Scheme of functioning of the program complex.***

Development of methods and algorithms of the software complex is caused by the following reasons:

- heterogeneity and many streaming user requests;
- non-stationarity of the parameters of the control object, limited information, uncertainty in the description of transient processes, the need for fuzzy structuring of ED;
- the need for tools for setting parameters in models describing transient processes in the

transmission of signal characteristics of ED frames.

The implemented complex of increasing the reliability of ED information is performed by running 10 program modules c_i, $i=1,\dots,10$,. For their testing, 2 conditional packages of tasks f_1 and f_2 are generated.

Estimates of the effectiveness of the complex are determined by the values of the following indicators in the EDS environment:

capacity: $c_1.v=100$, $c_2.v=...=c_5.v=50$, $c_6.v=...=c_{10}.v=10$;

productivity: $c_1=10$, $c_2=c_3=5$, $c_4=c_5=2$, $c_6=...=c_{10}=1$;

$c_i.H_0$ is the distribution of the processing time of the ED without errors is exponential with the parameter 0,01;

$c_1.H_1$ is the distribution of transmission time of ED with information distortions exponential with parameter 0,5;

$c_i.P_{01}$ is the distribution of detection and correction time in information;

$f_1.D_{in}$ is distribution of time interval between ED, respectively, without distortion of information, exponential, with parameter 0.7, and $f_2.D_{in}$ is distribution of time interval with errors in information, exponential, with parameter 0.3;

$f_1.D_{1t}$ and $f_2.D_{1t}$ are distribution of execution time of the job package of ED uniform on the segment [10, 30];

$f_1.D_{key}$ is distribution of time for tasks from the package f_1 with the task of monitoring the reliability of ED, which is distributed evenly;

$f_2.D_{key}$ is distribution of time for tasks from the package f_2, which is unknown beforehand and there is no the task of monitoring of information reliability;

$f_1.D_{len}$ and $f_2.D_{len}$ are distribution of the task execution length is uniform on the segment $[10^4, 2\cdot10^4]$;

$f_1.D_{prep}$ and $f_2.D_{prep}$ are distribution of preparation time for the task to start, has a fixed value equal 0,01;

D_{send} is distribution of transmission time of the packet is uniform over a segment $[10^{-3}, 2\cdot10^{-3}]$.

As a result of experimental calculations, the optimum values of parameters were obtained for tuning the following parameters of complex:

$g=(p.j).f$ is the basis for forming a package with length $p.l$;

$\tilde{z}$ is the sum $p.l$ of independent positive random summands with mean value $g\cdot\mu=1{,}5\cdot10^{-3}$ and dispersion $g.\sigma=10^{-4}$;

z is random variable with a normal distribution approximating the distribution $\tilde{z}$;

$p.t_{exec}=\max(0,z)$;

t_ε is the constant and equal 10^{-7};

$C_1=20$, $C_2=2$, $C_3=10^{-3}$, $C_4=10$ are coefficients of aimed function;

$p_{break}\in\{0{,}5;0{,}6;0{,}7;0.8;0{,}9;1\}$ is threshold of start interrupting;

$r_{run}\in R_{\max rs}, R_{\min rs}, R_{\max crs}, R_{\min csr}$ is the rating preference for setting the parameters of the software complex;

$R_{\max rs}(R_{\min rs})$ is the maximum (minimum) rating ;

$n_{run}\in\{0,....4\}$ is the number of simultaneously started modules;

$k^1_{check}\in\{0,...,4\}$ is the control time of the results of the complex;

$\Delta t_{dec}\in\{0{,}1;0{,}2;...;1\}$ is the interval between successive decisions on the reliability of ED information.

Conclusion

Thus, methodical bases for the implementation of program complex for ensuring the integrity, safety and reliability of the transmission and processing of ED information on the basis of the FSGN have been developed.

The methods of implementing the complex for importing data into the database for the subsequent processing of information using the methods of Data Mining have been developed. Methods for integrating the software modules of the complex in the FSGN environment are proposed.

The scope of the proposed methodology is not limited to a specific example of the problem of ensuring the reliability of ED transmission and processing. In the same way, optimizing methods of Data Mining and data processing of a wide variety of production systems can be constructed.

REFERENCES

1. Vendrov A.M. "*CASEtechnology. Modern methods and means of designing information systems"*, Finance and statistics. - Moscow, 1998, - p.176.

2. M.G.Konovalov, "*Methods of adaptive information processing and their applications"*, Moscow: IPI RAS, pp. 21-28, 2007.

3. I.I.Jumanov, A.R.Akhatov, *"Fuzzy semantic hyper network of information reliability control in electronic document management systems"*, 4th International Conference on Application of Information and Communication Technologies, Tashkent, 12-14 october pp. 21-25, 2010.

4. ARAkhatov, Kh.B.Karshiev, "*Control of the reliability of information of electronic documents on the basis of logical criteria and semantic redundancy Vestnik Scientific Researches"*, Samarkand: SamSU, no. 3, p. 55-62, 2014.

5. A.R.Akhatov, S.A.Tishlikov, *"Optimization of the reliability of transmission and processing of electronic documents on the basis of the properties of models of the transient process Chemical Technology"*, Control and management, TSTU, Tashkent, no. 1(61), p. 88-95, 2015.

6. F.L.George, *"Artificial intelligence, strategies and methods for solving complex problems"*, 4th edition., "New Mexico University", "Williams", MSANKPETERBURGKiev: pp. 18-63, 2005.