Using Fuzzy Logic in the Research of the Data Visualization Process in Infographics

Oleksandr Tymchenko[a,b], Orest Khamula[b], Svitlana Vasiuta[b], Olha Lukovska[b], Solomiya Dorosh[b] and Olha Sosnovska[b]

[a] *University of Warmia and Mazury, Olsztyn, Poland*
[b] *Ukrainian Academy of Printing, Lviv, Ukraine*

Abstract
Data visualization in infographics is a rather interesting and yet little-studied process that requires comprehensive research. In the study, an analysis of the factors identified by an expert survey and divided by importance, which affect the visualization of data in infographics, is carried out. The analytic hierarchy process and the ranking method were used for the research. These methods show good calculation results when studying processes in which factors cannot be described using mathematical data. But, in the vast majority, the results obtained by these two methods give different results. Therefore, to verify the precision of the obtained results, the researchers recommend using the calculation of integral indicators of the quality of the considered process based on fuzzy logic.
To calculate the integral quality indicators, it is required to establish a universal set of terms of values of isolated linguistic variables. As a result, this solution lies in the description of linguistic variables with the assignment of their designations, determination of recommended limits of values of universal fuzzy sets, and linguistic terms for the comparative process, construction, and calculation of functions belonging to linguistic variables.

Keywords [1]
Influence factors, infographics, data visualization, linguistic variables, fuzzy sets, fuzzy logic, membership function, integral quality indicator

1. Introduction

It is a well-known fact that the key way of obtaining information about the world is through vision. On average, this number is 90%. Perceived information allows to memorize information faster and, accordingly, to increase labor productivity. Information presented in the form of pictures, diagrams, or in any other presentation of infographics, is even better perceived and memorized. This is due to the fact that the information presented in the form of infographics allows to better visualize certain details and draw attention to specific information better than text.

Usually, data visualization covers large amounts of information and condenses it. Another purpose of visualization is that it makes the perception of complex information more accessible and allows you to create parities in the comparison of quantities.

A crucial task of visualization is its persuasiveness. Therefore, distortion of the presented information should be avoided. Correctly presented information does not interfere with the perception of details.

That is why, when creating a visualization, it is important to consider the effectiveness of the data presented, what is worth sharing, and presenting it in a way that engages the user. The next step of

SCIA-2022: 1st International Workshop on Social Communication and Information Activity in Digital Humanities, October 20, 2022, Lviv, Ukraine
EMAIL: olexandr.tymchenko@uwm.edu.pl (O. Tymchenko); khamula@gmail.com (O. Khamula); svitlanavasyta@gmail.com (S. Vasiuta); olhalukovska@yahoo.com (O. Lukovska); solomiadorosh@gmail.com (S. Dorosh); olhakh@gmail.com (O. Sosnovska)
ORCID: 0000-0001-6315-9375 (O. Tymchenko); 0000-0003-0926-9156 (O. Khamula); 0000-0003-0078-9740 (S. Vasiuta); 0000-0002-4074-7454 (O. Lukovska); 0000-0002-6824-5421 (S. Dorosh); 0000-0001-5413-2517 (O. Sosnovska)

this process is choosing the method of creating the visualization itself, which will convey specific information. Therefore, the study of various visualization techniques today is quite an essential issue and deserves the attention of researchers.

2. Related works

There is a sufficient number of publications and discussions regarding the research of data visualization processes and the use of infographics in general in practice. This, in our opinion, indicates that this topic is quite relevant.

Among the publications that reveal research on this topic, we would like to draw attention to the following. One of the main ways of using infographics, especially for scientists and teachers, is in the field of education. Among the publications that reveal research on this topic, we would like to draw attention to the following. One of the main ways of using infographics, especially for scientists and teachers, is in the field of education. It is necessary to provide the essential information concisely and clearly so that in a short time, it is possible to present a maximum of information and it is understandable for everyone. Thus, in study [1], the author considers the structure issue of the presentation of information visualization elements. In this work, the author notes that the study of the information visualization process is quite a significant topic for teachers, as they try to understand and help students gain new knowledge with the help of modern literature and practices. The author highlighted the following elements of visualization: text, images, data, and the interaction between them, noting and exploring their impact on the audience.

In research [2], the authors reveal another field of data visualization — the military. The authors indicate how it is necessary to correctly create a visualization of the required data so that they can be quickly comprehended, perceived, and an accurate decision could be made. In this study, the most modern data transmission methods, and the main tasks of visualization are considered. As a result of the research, the authors proposed a concept demonstrator for visualizing collaborative effort modeling.

Important research in the field of data visualization or information in general is the work of authors [3], which focuses on the suitability of specific color combination types. The work distinguishes seven types: harmonious colors, opposite colors, highly saturated colors, low saturated colors, high-brightness, low-brightness colors, and disharmonious colors. Evaluation of these studies focused on color visibility. Studies have shown that preference is still given to harmonious colors, contradicting previous claims that disharmonious colors can be beneficial.

Another vital field of data visualization is management. The authors proposed options and technology for processing the text array and visualization of it in the study [4With the help of the proposed technology, it will be possible to highlight valuable information hidden in a dense abstract form of text and transform it into simple and intuitive information. This, in turn, will enable managers to quickly understand and make more valuable decisions. As the questionnaire regarding this technology showed, the proposed technology facilitates quick and informative access to key information, which will ultimately reduce the workload and time needed to realize the status of their projects.

Cognitive forms of the human interaction process with visualization are considered in work [5]. The authors discovered when there is a cognitive understanding between the type of problem being solved and the information, the task load index is lower, compared with the case when there is no cognitive fit. The research data confirm the relevance of the study of the data visualization process and its effectiveness. Also, the research data are crucial because they provide greater validity to cognitive theory and assessment of cognitive load, as well as effort in general.

Every day, the amount of data that needs to be stored and processed increases. Accordingly, the need for tools that can provide the correct visualization of data has increased, which, in turn, has led to the development of various programming software for data visualization. But the use of these software tools requires users to develop visualization from scratch. An alternative is to create interactive visualization design environments. Interesting examples of such systems are NarVis [6], TechSpectogram [7], and VisComposer [8]. These systems allow users to create a variety of visualizations with the possibility of previewing and optimizing the data in real-time. In work [9], the authors considered the issue of automated generation of visualization, tools, and documents of recommendations for network graphs usage. This study is also important in terms of the challenges and promising directions for further research in the field of automated infographics and visualization recommendations.

In conclusion, it is important to note the study [10] of authors who consider the principles of effective data visualization. The paper states that researchers do not always present information correctly or use incorrect data visualization practices. The authors provide ten principles that serve to improve data visualization, including technical aspects and, for example, the combination of colors. All these principles and proposals are designed for the most effective presentation of specific data.

3. The methodology of using fuzzy logic in researching the process of data visualization in infographics

With the help of previous research [11], using the analysis of expert judgments, we established a certain list of factors that affect the process of data visualization in infographics. From this list, six of the most significant in this process were chosen, namely: text (T), numerical data (ND), graphs and charts (GC), flowcharts (FC), image (IM), and icons (IC). The next question that was considered in this paper is how these factors influence each other. A dependency graph of the relationships between factors was also constructed. In work [12], we investigated which factor is more important in the process of data visualization. For this, the analytic hierarchy process was used. As a result of the research, we obtained an optimized model of influencing factors in the data visualization process in infographics. The factors were divided into 5 levels. From this model, we can conclude that the most important factor in the data visualization process was the use of icons. Our next step was a study [13], namely a comparison of different methods of determining the priority or importance of a specific factor with others. A study of the priority of influencing factors in the data visualization process was conducted using two methods analytic hierarchy process and the ranking method. As research has shown — the ranking method appeared to be the best in this case. It showed the distribution of factors in 6 priority levels. This method is more comprehensive since it takes into account indirect dependencies that prefer one factor over another.

Since the data visualization process is described by factors that are difficult to represent in the form of specific mathematical units, it is advisable to use fuzzy logic to clarify the accuracy of the results. It became necessary to use a mathematical tool with the help of which ambiguous expert statements would be translated into the language of clear mathematical formulas. As research shows, the usage of fuzzy logic is a good solution as it provides an opportunity to operate the fuzzy input data. In general, fuzzy logic uses linguistic variables with the help of explicit rules.

Fuzzification is the basis of the use of fuzzy logic. Its task is to transform the initial data into a set that would correspond to the terms of the linguistic variable. These data can be described by one or several terms. The degree of their correspondence to the term is given as the degree of membership on a fuzzy set. As it follows from using the theory of fuzzy logic, fuzzification is the base for evaluating the quality of the process under research. For calculations, fuzzy logical equations and a knowledge base are used. The knowledge base is built according to expert data that are based on fuzzy linguistic rules with the condition – "if-then." Accordingly, the operation that involves mapping the obtained numerical results of the calculation is called defuzzification. Moreover, these numerical results may have different units of measurement.

Fuzzy sets and linguistic variables are the basis of any fuzzy logic, which, in turn, are based on membership functions constructed using a set of terms and linguistic terms of the factors under study.

The main values of a linguistic variable are words and phrases in natural or formal expression. Of course, the data presented in the form of numbers is more precise than the data presented in words, regarding with which the linguistic variable describes the phenomena approximately, because they are complex and cannot be described in quantitative terms. It can be stated that the purpose of fuzzy logic is the same as words or phrases in living language. When comparing linguistics and fuzzy variables, it can be noted that linguistic variable are of a higher order. Linguistic variables include a number of rules that make up the majority of the description of its structure, namely syntactic: can be set grammatically; can specify an algorithmic procedure for calculating the meaning of the values.

Conducting calculations using the theory of fuzzy logic involves performing some operations divided into certain stages and, accordingly, calculations presented below.

A. The model of forming an integral indicator of process quality

In order to better understand the complex process, the method of fuzzy logic is recommended to divide into certain sub-processes, which could be better described [14]. The data visualization design process in infographics, which is considered in the research, will be denoted by, for example, the function G, which in turn can be divided into two subprocesses: the argument Q, which determines the quality of the artistic presentation of information, and the argument V, which defines the technical features quality of this process. The arguments of the functions are factors. In our calculations, these are linguistic variables obtained and described in previous studies [12]. The values of the separated functions form integral quality indicators, which are determined, in turn, by partial quality indicators of linguistic variables, which are divided into groups according to their functional purpose.

Thus, using the theory of fuzzy logic, some function G determines the quality of data visualization in infographics, meaning:

$$G = F_G(Q, V). \tag{1}$$

Accordingly, the arguments of this function can be described:

- argument Q, which determines the quality of the first subprocess:

$$Q = F_q(q_1, q_2, q_3), \tag{2}$$

where q_i – linguistic variables of the argument Q.

- argument V, which determines the quality of the second subprocess:

$$V = F_v(v_1, v_2, v_3), \tag{3}$$

where v_i – linguistic variables of the argument V.

For further research, we need to establish a universal set of terms of values of separated linguistic variables, so it, in turn, corresponds to the limits of the existence of linguistic variables. As a result, this solution is in the form of a table with linguistic variables, specifying their designations, and determining the recommended limits of values of universal sets of terms and linguistic terms for the comparative research process of data visualization in infographics.

Based on the data obtained by this method, we can start building a multi-level model for the formation of integral process quality indicators.

The use of a multi-level model of fuzzy logical derivation contributes to the consistent establishment of a forecast of the implementation quality of the publication structuring process by accumulating knowledge from the lowest to the highest of its levels.

B. Construction and calculation of membership functions of linguistic variables

The next stage of solving the given task according to the theory of fuzzy logic is the construction and calculation of the membership functions of linguistic variables [15]. The membership function is the main characteristic of the formation level of the process quality, which in turn is given by linguistic terms, such as *N* and *G*.

In the calculations, it is assumed that the original database is a universal fuzzy set, for example, *B*, which should be divided into certain parts (quanta). At these points of division, we set the separated linguistic variables and ranks, with the help of which linguistic terms can be identified. Let's set our universal fuzzy set as $B = \{b_1, b_2, \ldots, b_n\}$, and also set the corresponding ranks $r_N(b_i)$ – for the study of platforms for the implementation of distance learning and $r_J(b_i)$ – visualization of data in infographics in the ranges of b_i ($i = 1, \ldots, n$).

Then, taking into account the above assumptions, the linguistic term for our researched process, which we will call *Y*, can be represented in the form of a certain fuzzy set:

$$Y_F = \left\{\frac{\mu_y(b_1)}{b_1}, \frac{\mu_y(b_2)}{b_2}, \frac{\mu_y(b_3)}{b_3}\right\}, \tag{4}$$

where $Y_F \subset B$; $\mu_y(b_i)$ – degree of mentorship to the set Y_F of the element $b_i \in$ B.

The membership functions, or the membership degree, which are specified as the value $\mu_y(b_i)$, become the basis of logical expressions for the numerical expression of the linguistic term *Y*. The distribution of which is given in the form:

$$\frac{\mu_1}{r_1} = \frac{\mu_2}{r_2} = \cdots = \frac{\mu_n}{r_n}, \tag{5}$$

where $\mu_i = \mu_y(b_i)$ – degree of mentorship; $r_i = r_y(b_i)$ – corresponding ranks for all $i = 1, \ldots, n$.

The normalization condition must also be fulfilled for the membership functions, namely their sum must be equal to 1:

$$\mu_1 + \mu_2 + \cdots + \mu_n = 1, \tag{6}$$

To determine the ranks of factors established by experts or calculated using ranking methods or the analytic hierarchy process, their numerical values can be obtained using the following ratios:

$$\left.\begin{array}{l} \mu_1 = \left(1 + \frac{r_2}{r_1} + \frac{r_3}{r_1} + \cdots + \frac{r_n}{r_1}\right)^{-1}; \\ \mu_2 = \left(\frac{r_1}{r_2} + 1 + \frac{r_3}{r_2} + \cdots + \frac{r_n}{r_2}\right)^{-1}; \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots; \\ \mu_n = \left(\frac{r_1}{r_n} + \frac{r_2}{r_n} + \frac{r_3}{r_n} + \cdots + 1\right)^{-1}. \end{array}\right\}, \tag{7}$$

For a better visual perception of linguistic terms and their high-quality graphic display, the range of values of linguistic variables specified in the relevant tables discussed above is presented in the form of three points. As a result, we will get a square inverse symmetric matrix $X = x_{ij}$, where $x_{ij} = r_i / r_j$ for $i, j = 1, 2, 3$, and the estimates of the ranks of linguistic terms should also be considered.

The result of our task to obtain an integral indicator of the quality level of the process researched is presented as follows:

$$\left.\begin{array}{l} G_F = F(j_x, v_e) \to max, x = \overline{1,3}; e = \overline{1,3} \\ j_x > 0, \;\; v_e > 0, \\ \mu_N(b_i) \to max, b_i \in B, G_F \subset \mathrm{B}, \mathrm{i} = \overline{1,3} \end{array}\right\}. \tag{8}$$

The goal of the research, according to formula (8), is to achieve the maximum value of the functions that characterize the level of quality of the considered process. And the next steps of calculations according to this method are the construction of a matrix of pairwise comparisons for all linguistic terms, using the scale of the relative importance of objects. In this case, the components of the eigenvector of the matrix will be the ranks of the linguistic terms, which in turn are used to calculate the values of the membership functions μ_i for each of the terms Therefore, the final numerical values of the membership functions at the determined ranks of the linguistic terms at the three points of division of the universal set (b_1, b_2, …, b_n) will be obtained as a result of the matrix calculation:

$$U = \begin{bmatrix} 1 & \frac{r_2}{r_1} & \frac{r_3}{r_1} \\ \frac{r_1}{r_2} & 1 & \frac{r_3}{r_2} \\ \frac{r_1}{r_3} & \frac{r_2}{r_3} & 1 \end{bmatrix} \tag{9}$$

C. Designing a fuzzy knowledge base and a system of fuzzy logic equations

For the next studies, namely the study of the relationships between the separated factors of influence on the relevant processes, a fuzzy knowledge base should be involved. It represents a set of fuzzy rules of the "if-then" type and makes it possible to link information between the input and output object researched. This implementation is possible when using fuzzy sets and linguistic variables [16]. This combination allows obtaining dependencies between physically separated processes, as already mentioned above.

To determine the procedure for obtaining the degree of the membership of functions to a set of linguistic terms of the quality indicator of the researched processes, an established fuzzy knowledge base is used in the form of a system of fuzzy logic equations.

The basis for considering a fuzzy knowledge base is expert judgment of the impact of factors on the considered processes. Also, this base implements an algorithm, with the help of which the forecasting of the quality of processes is formed, taking into account the obtained combinations of the values of the relevant linguistic terms.

At this stage, there is a need to build a fuzzy knowledge base, as well as to calculate an integral indicator of the quality of the process of data visualization in infographics as a function of the highest level and defined by the linguistic variable G with its linguistic terms "low," "medium," and "high."

And the functions of the lower level are sets of linguistic variables Q and V, which in turn are also represented by the linguistic terms "low," "medium," and "high."

According to the fuzzy knowledge base, taking into account equality (1) and based on the obtained model, we receive the following general form:

IF (Q = low) I (Q = medium) I (Q = high)
I (V = low) I (V= medium) I (V = high)
THEN (G = low) I (G = medium) I (G = high)

Based on the above information, we construct the linguistic variable G table of the knowledge matrix.

Our next step should be to obtain the membership function values for our sets of terms. We can solve this task by forming fuzzy logic equations based on the obtained knowledge matrices G. In this way, we will be able to obtain an integral indicator of the quality of the process researched. Accordingly, we obtain the following equalities of the membership function of the linguistic variable G:

- for the term "low," the following is received
  $\mu_{low}(G) = \mu_{low}(Q) \wedge \mu_{low}(V) \vee \mu_{low}(Q) \wedge \mu_{medium}(V)$;
- for the term "medium," the following is received
  $\mu_{medium}(G) = \mu_{medium}(Q) \wedge \mu_{low}(V) \vee \mu_{medium}(Q) \wedge \mu_{medium}V)$;
- for the term "high," the following is received
  $\mu_{high}(G) = \mu_{high}(Q) \wedge \mu_{medium}(V) \vee \mu_{high}(Q) \wedge \mu_{high}(V)$.

Using the same technique, we obtain fuzzy knowledge bases and linguistic equations for the lower level of the corresponding linguistic variables, for example, $Q – Y(q_1, q_2, q_3)$ and $V – Y(v_1, v_2, v_3)$, which belong to the process under study.

Based on the obtained generalized variants of the logical expression of linguistic variables Q and V, let's build knowledge matrices.

Based on the obtained knowledge matrices and using a system of logical equations, the membership functions for our linguistic variables Q and V are obtained. Accordingly, the membership functions for the linguistic variable Q can be represented as follows:

- for the term "low"
  $\mu_{low}(Q) = \mu_{simple}(q_1) \wedge \mu_{complicated}(q_2) \wedge \mu_{simple}(q_3)$
  $\vee \mu_{complicated}(q_1) \wedge \mu_{simple}(q_2) \wedge \mu_{simple}(q_3)$
- for the term "medium"
  $\mu_{medium}(Q) = \mu_{simple}(q_1) \wedge \mu_{simple}(q_2) \wedge \mu_{complicated}(q_3)$
  $\vee \mu_{complicated}(q_1) \wedge \mu_{complicated}(q_2) \wedge \mu_{simple}(q_3)$
- for the term "high"
  $\mu_{high}(Q) = \mu_{complex}(q_1) \wedge \mu_{complex}(q_2) \wedge \mu_{complicated}(q_3)$
  $\vee \mu_{complicated}(q_1) \wedge \mu_{complicated}(q_2) \wedge \mu_{complex}(q_3)$.

Accordingly, the membership functions for the linguistic variable V can be represented as follows:

- for the term "low"
  $\mu_{low}(V) = \mu_{simple}(v_1) \wedge \mu_{small}(v_2) \wedge \mu_{simple}(v_3)$
  $\vee \mu_{simple}(v_1) \wedge \mu_{medium}(v_2) \wedge \mu_{simple}(v_3)$
- for the term "medium"
  $\mu_{medium}(V) = \mu_{simple}(v_1) \wedge \mu_{small}(v_2) \wedge \mu_{complex}(v_3)$
  $\vee \mu_{simple}(v_1) \wedge \mu_{small}(v_2) \wedge \mu_{complicated}(v_3)$
- for the term "high"
  $\mu_{high}(V) = \mu_{complex}(v_1) \wedge \mu_{large}(v_2) \wedge \mu_{complicated}(v_3)$
  $\vee \mu_{complex}(v_1) \wedge \mu_{medium}v_2) \wedge \mu_{simple}(v_3)$.

Based on the obtained equalities, we form fuzzy sets of linguistic variables Q and V with corresponding membership functions for the accepted terms "low," "medium," and "high" in the form of some fuzzy set:

$$G(Q,V) = \left\{ \frac{\mu_{\mathrm{low}}(G)}{n_1}, \frac{\mu_{medium}(G)}{n_2}, \frac{\mu_{high}(G)}{n_3} \right\}, \quad (10)$$

where n_1, n_2, n_3 – quantitative values of the linguistic variable G regarding the defined terms.

The defuzzification is performed based on the expert knowledge base and using the knowledge matrices of the linguistic variable G. In reference sources, defuzzification is understood as a certain procedure of transforming a fuzzy set into a clear number by the degree of membership. Its purpose is to obtain the usual quantitative value of each of the initial variables, external concerning the fuzzy inference system, by using the results of the accumulation of all the initial linguistic variables. For this, the pre-numbered values of the membership functions at the three points of division of the selected universal set *A* for all linguistic variables are presented.

To obtain the real weightings of the membership functions of the linguistic variables *Q* and *V*, the values of the defined terms "low," "medium," and "high" should be substituted into fuzzy logic equations. It should be noted that the membership functions of the linguistic variables are calculated based on the initial values of the linguistic variables presented by the experts. After obtaining the values of the membership functions of the lower-level linguistic variables and taking the above formulas as a basis, the membership function of the highest-level linguistic variable G is calculated according to the accepted terms "low", "medium", and "high".

In the final stage of calculations according to this method, namely, using the defuzzification of fuzzy sets and the center of the mass method, the indicators of the level of the predicted quality of process implementation are calculated. According to this, we receive:

$$G = \frac{\sum_{i=1}^{n} \left[\underline{G} + (i-1) \frac{\overline{G} - \underline{G}}{n-1} \right] \mu_i(G)}{\sum_{i=1}^{n} \mu_i(G)} \tag{11}$$

where $\underline{G}, \overline{G}$ – minimum and maximum values of quality level indicators; *n* is the number of specified fuzzy terms (in our case, 3).

The values of other variables present in expression (9) are taken as follows: $\mu_1(G) = \mu_{low}(G)$; $\mu_2(G) = \mu_{medium}(G)$; $\mu_3(G) = \mu_{high}(G)$; values of the lower and higher limits for the linguistic variable G and are taken as equal, respectively – $\underline{G}$ = 5%, $\overline{G}$ = 100%. It is recommended to calculate indicators of the predicted quality level of the process at three points of the interval from 5 to 100%. The predicted quality should be greater than 50%. Then it is considered that the conducted research is accurate.

4. Research results

Taking this method as a basis, let's calculate the indicator of the predicted level of the data visualization process quality in infographics. The process of designing data visualization in infographics will be denoted by function *G*. The arguments of the functions are factors of linguistic variables. Integral quality indicators are formed by function values and are determined by partial quality indicators of linguistic variables, divided into groups according to their functional purpose. This process of designing data visualization in infographics will be divided into two sub-processes, namely: *Q*, which determines the quality of the artistic presentation of information, and *V*, which determines the quality of technical features in this process.

Taking expressions (2) and (3) into account, we will describe the linguistic variables for these attributes.

- for the *Q* argument: q_1 – "data visualization type;" q_2 – "color gamut of visualization;" q_3 – "special effects;"
- for the *V* argument: v_1 – "labels in visualization;" v_2 – "scale of elements in visualization;" v_3 – "presenting data in visualization."

We establish a universal set of terms of values of isolated linguistic variables so that it, in turn, corresponds to the limits of the existence of linguistic variables. As a result, this solution is shown in Table 1 with linguistic variables, specifying their designations and determining the recommended limits of values of a universal set of terms and linguistic terms for the comparative research process of data visualization in infographics.

Table 1
Set of terms of linguistic variables values

Variable	Linguistic essence of the variable	Universal set of values (set A)	Linguistic terms (set B)
q1	type of data visualization (types of infographics)	(1-9) CU	Simple, complicated, complex
q2	color gamut of visualization	(1-3) CU	RGB, CMYK, HSB
q3	special effects	(1-9) CU	Simple, complicated, complex
v1	inscriptions in visualization	(1-9) CU	Simple, complicated, complex
v2	scale of elements in the visualization	(10-95) %	Small, medium, large
v3	data presentation in visualization	(1-9) CU	Simple, complicated, complex

In this table, the linguistic variable "type of data visualization" is presented as a variety of infographics; the linguistic variable "data presentation in visualization" identifies the complexity of the presentation; linguistic variable "scale of elements in the visualization" – the geometric dimensions in percentage. For universal sets of values, where it is impossible to give a numerical explanation, the scale of relative importance of objects, according to which there are 9 importance ratings, is taken into account.

Due to this, their conditional values for q_1, q_3, v_1 and v_3 are taken to be equal to (1-9) CU (Conventional Units), and for q_2 equal to (1-3) CU.

Taking the obtained data as a basis, we begin to build a multi-level model of the formation of integral indicators of the data visualization quality in infographics (Fig. 1).

The use of a multi-level model of fuzzy logical derivation contributes to the consistent establishment of a forecast of the implementation of the publication structuring process quality by accumulating knowledge from the lowest to the highest of its levels. This multi-level model includes subordinate models: a model of the quality of artistic presentation of information, a model of the quality of technical features.

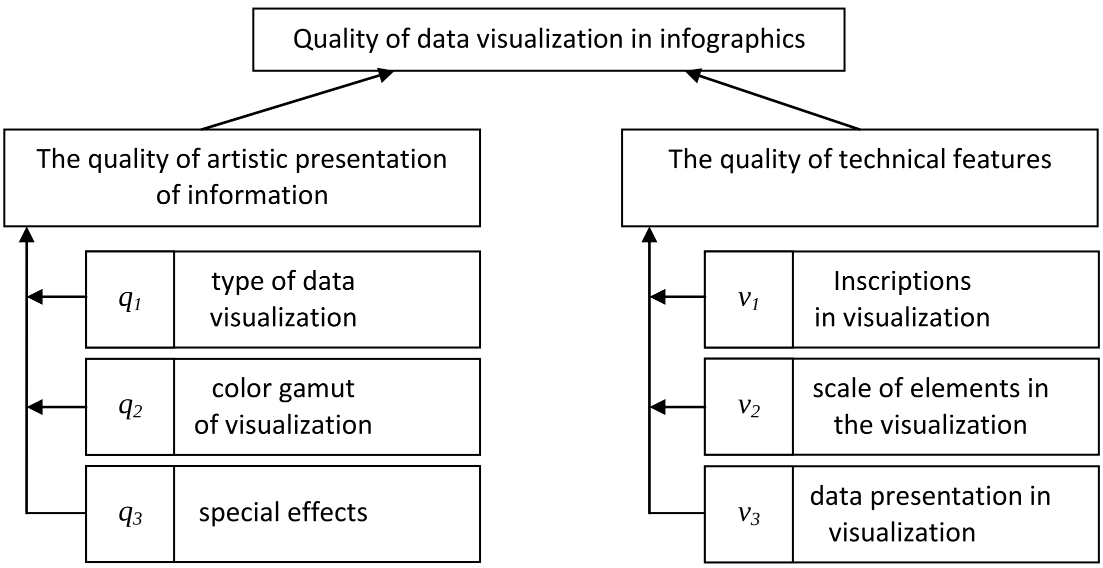

Figure 1: A multi-level model of the formation of integral indicators of the data visualization quality in infographics

The resulting model reflects the logic of forming the quality of the data visualization process in infographics with the linguistic variables identified in the survey process, according to expressions (1, 2, 3), which become a source for calculating integral quality indicators.

Let's calculate the membership functions to the linguistic quality variables of the data visualization process in infographics. The resulting model (Fig. 1) reflects the logic of forming the quality of the data visualization process in infographics with the linguistic variables identified during the survey.

Using matrix (9) as a basis, we construct matrix *Y*. For the linguistic variable q_1 "type of visualization," the universal set of values is equal to A (q_1) = [1; 9] CU, and the set of terms of values for it is equal to B (q_1) = < *simple*, *complicated*, *complex* >. According to expert judgments and according to the data in Table 1, the universal set has the following values at the points of division: $a_1 = 1$; $a_2 = 5$; $a_3 = 9$. The appearance of the matrices for the corresponding terms "simple," "complicated," and "complex" will have the following form:

$$Y_{simple}(q_1) = \begin{bmatrix} 1 & \frac{7}{9} & \frac{1}{9} \\ \frac{9}{7} & 1 & \frac{1}{7} \\ 9 & 7 & 1 \end{bmatrix}, Y_{complicated}(q_1) = \begin{bmatrix} 1 & 9 & 1 \\ \frac{1}{9} & 1 & \frac{1}{9} \\ 1 & 9 & 1 \end{bmatrix}, Y_{complex}(q_1) = \begin{bmatrix} 1 & 7 & 9 \\ \frac{1}{7} & 1 & \frac{9}{7} \\ \frac{1}{9} & \frac{7}{9} & 1 \end{bmatrix}$$

The last line of the term *"simple"* is determined by an expert method and the rank of the variable decreases. Using the values of the matrix and relation (9), as well as using the program for calculating the method of binary comparisons, we obtain the values of the membership functions for the terms *"simple," "complicated,"* and *"complex"*:

$\mu_{simple}(a_1) = 0{,}53$; $\mu_{complicated}(a_1) = 0{,}09$; $\mu_{complex}(a_1) = 0{,}059$;
$\mu_{simple}(a_2) = 0{,}412$ $\mu_{complicated}(a_2) = 0{,}82$; $\mu_{complex}(a_2) = 0{,}\ 412$;
$\mu_{simple}(a_3) = 0{,}059$; $\mu_{complicated}(a_3) = 0{,}09$; $\mu_{complex}(a_3) = 0{,}53$.

The next step is to perform the normalization of the values of the membership functions, taking into account the established linguistic terms relative to the unit, and obtain the normalization coefficient according to the formula:

$$k_n = \frac{1}{\max \mu_n\,(a_i)}, (i = 1, 2, 3), \tag{12}$$

where n– terms of the corresponding linguistic variable; $\mu_{nj}(a_i) = k_n \times \mu_n(a_i)$.

Let's normalize the values of the membership functions of the linguistic variable "communication tools," taking into account the above expression:

$\mu_{simplej}(a_1) = 1{,}00$; $\mu_{complicatedj}(a_1) = 0{,}11$; $\mu_{complexj}(a_1) = 0{,}11$;
$\mu_{simplej}(a_2) = 0{,}77$; $\mu_{complicatedj}(a_2) = 1{,}00$; $\mu_{complexj}(a_2) = 0{,}77$;
$\mu_{simplej}(a_3) = 0{,}11$; $\mu_{complicatedj}(a_3) = 0{,}11$; $\mu_{complexj}(a_3) = 1{,}00$;

In order to display linguistic terms in fuzzy sets, it is necessary to use the normalized values of the membership functions of the linguistic variable "type of visualization" (types of infographics) and equality (4). Accordingly, we receive:

simple types of infographics $= \left\{\frac{1}{1}, \frac{0{,}77}{5}, \frac{0{,}11}{9}\right\}$ *CU*;

complicated types of infographics = $\left\{\frac{0{,}11}{1}, \frac{1}{5}, \frac{0{,}11}{9}\right\}$ CU;

complex types of infographics = $\left\{\frac{0{,}11}{1}, \frac{0{,}77}{5}, \frac{1}{9}\right\}$ CU.

The data obtained in the calculation process of the membership functions of linguistic variables become the basis for constructing a fuzzy knowledge base and a system of fuzzy logical equations. On their basis, the design and calculation of the integral indicator of the quality level of research processes of platforms for the implementation of distance learning and data visualization in infographics are carried out.

The next stage of the considered process in the research is the need to build a fuzzy knowledge base, as well as to calculate an integral indicator of the quality of the data visualization process in infographics as a function of the highest level and defined by the linguistic variable G – "data visualization quality in infographics" and its linguistic terms "low", "medium", and "high." As well as the functions of the lower level – the set of linguistic variables Q – "quality of artistic presentation" and V – "quality of technical presentation", that in turn are also represented by the linguistic terms "low", "medium", and "high".

According to the theory of the fuzzy knowledge base, taking into account equality (1) and based on the obtained model from (Fig. 1), the following general view is obtained for data visualization in infographics "if - then" is presented on page 6.

Based on the above information, let's start building the linguistic variable G Table 2 of the knowledge matrix:

Table 2

Table of the knowledge matrix of the linguistic variable G

The quality of artistic representation (Q)	The quality of technical representation (V)	The quality of data visualization in infographics (G)
low	low	low
low	medium	
medium	low	medium
medium	medium	
high	medium	high
high	high	

Using the same technique, we obtain fuzzy knowledge bases and linguistic equations for the lower level of the corresponding linguistic variables $Q - Y(q_1, q_2, q_3)$, $V - Y(v_1, v_2, v_3)$, which refer to the process of data visualization in infographics.

Based on the obtained equalities, let's form fuzzy sets of linguistic variables Q and V with corresponding membership functions for the accepted terms "low," "medium," and "high" in the form of some fuzzy set (10).

The defuzzification is performed based on the expert knowledge base and using the knowledge matrices of the linguistic variable G given in table (2) and the given fuzzy sets (10). Its purpose is to obtain the usual quantitative value of each of the initial variables, external concerning the fuzzy inference system, by using the results of the accumulation of all the initial linguistic variables. For this, the pre-numbered values of the membership functions at the three points of division of the selected universal set A for all linguistic variables are presented.

To obtain the real weightings of the membership functions of the linguistic variables Q "quality of artistic presentation" and V "quality of technical presentation," the values of the defined terms "low," "medium," and "high" should be substituted into fuzzy logic equations. It should be noted that the membership functions of the linguistic variables are calculated based on the initial values of the linguistic variables presented by the experts.

To calculate the linguistic variables Q, the values of the corresponding terms should be taken from the earlier calculations into fuzzy logic equations, and we will receive:

- for the term "low"
  $\mu_{low}(Q) = 0,77 \wedge 1 \wedge 0,55 \vee 1 \wedge 0,77 \wedge 0,55 = 0,55$
- for the term "medium"
  $\mu_{medium}(Q) = 0,77 \wedge 0,77 \wedge 1 \vee 1 \wedge 1 \wedge 0,55 = 0,77$
- for the term "high"
  $\mu_{high}(Q) = 0,77 \wedge 0,66 \wedge 1 \vee 1 \wedge 1 \wedge 0,66 = 0,66.$

Accordingly, the membership functions for the linguistic variable V can be represented as follows:

- for the term "low"
  $\mu_{low}(V) = 0,77 \wedge 0,77 \wedge 0,66 \vee 0,77 \wedge 1 \wedge 0,66 = 0,66$
- for the term "medium"
  $\mu_{medium}(V) = 0,77 \wedge 0,77 \wedge 0,55 \vee 0,77 \wedge 0,77 \wedge 1 = 0,77$
- for the term "high"
  $\mu_{high}(V) = 0,33 \wedge 0,45 \wedge 1 \vee 0,33 \wedge 1 \wedge 0,66 = 0,45.$

By obtaining the values of the membership functions of the lower-level linguistic variables and using the above formulas as a basis, let's proceed to calculate the membership function of the highest-level linguistic variable G "data visualization quality in infographics" according to the accepted terms "low," "medium," "high."

Thus, the following values are received:

- for the term "low"

  $\mu_{low}(G) = 0{,}55 \wedge 0{,}66 \vee 0{,}55 \wedge 0{,}77 = 0{,}55$
- for the term "medium"

  $\mu_{medium}(G) = 0{,}77 \wedge 0{,}66 \vee 0{,}77 \wedge 0{,}77 = 0{,}77$
- for the term "high"

  $\mu_{high}(G) = 0{,}66 \wedge 0{,}77 \vee 0{,}66 \wedge 0{,}45 = 0{,}66$.

In the final stage of calculations according to this method, namely, using the defuzzification of fuzzy sets (10) and the center of the mass method, we calculate the indicators of the level of the predicted quality of the implementation of the data visualization process in infographics according to expression (11). The calculation of the indicator of the predicted level of the data visualization process quality in the infographics is carried out at three points of the interval – 5, 55, and 100%, and the following numerical value is received:

$$G = \frac{5 \times 0{,}55 + 55 \times 0{,}77 + 100 \times 0{,}66}{0{,}55 + 0{,}77 + 0{,}66} = 52{,}5\%$$

5. Conclusions

From the material described above, it is clear that the data visualization process is described by factors that are difficult to represent in the form of certain mathematical units. Therefore, to clarify the accuracy of the results of expert judgment, it is advisable to use such methods as the theory of fuzzy logic. When analyzing the previous results, namely using the analytic hierarchy process and the ranking method, it became necessary to use a mathematical tool with the help of which the ambiguous statements of experts would be translated into the language of clear mathematical formulas. As research shows, the usage of fuzzy logic is a convenient solution. It provides an opportunity to operate the fuzzy input data.

Based on the calculations, using the theory of fuzzy logic, an indicator of the predicted level of quality of the data visualization process in infographics was established. As the results showed, it was 52.5%, which is greater than the value of 50%, which means that our assumptions and calculations were correct.

Therefore, using the theory of fuzzy logic helps to provide a qualitative assessment of the process researched based on the calculation of the indicator of its level of predicted quality.

6. Acknowledgements

The authors of this study are sincerely grateful to the large audience of students and teachers who contributed to the research and establishment of influencing factors on the data visualization process in infographics. Special gratitude to colleagues for help and relevant suggestions that helped to improve the submitted material.

7. References

[1] M. Sorapure, Text, Image, Data, Interaction: Understanding Information Visualization**.** Computers and Composition, Volume 54, December 2019, 10251.

[2] H. Shen, T. Bednarz, H. Nguyen, F. Feng, T. Wyeld, P. J. Hoek, E. H.S. Lo, Information visualisation methods and techniques: State-of-the-art and future directions, Journal of Industrial Information Integration, Volume 16, December 2019, 100102.

[3] S. Einakian, T. S. Newman, An examination of color theories in map-based information visualization, Journal of Computer Languages, Volume 51, April 2019, pp. 143-153.

[4] J. Sun, K. Lei, L. Cao, B. Zhong, Y. Wei, J. Li, Z. Yang, Text visualization for construction document information management, Automation in Construction, Volume 111, March 2020, 103048.
[5] J. K. Nuamah, Y. Seong, S. Jiang, E. Park, D. Mountjoy, Evaluating effectiveness of information visualizations using cognitive fit theory: A neuroergonomics approach, Applied Ergonomics, Volume 88, October 2020, 103173.
[6] C. Edmond, T. Bednarz, Three trajectories for narrative visualization, Visual Informatics, Volume 5 (2021), pp. 26–40.
[7] E. Perez-Molina, F. Loizides, Novel data structure and visualization tool for studying technology evolution based on patent information: The DTFootprint and the TechSpectrogram, World Patent Information, Volume 64, March 2021, 102009.
[8] H. Mei. W. Chen, Y. Ma, H. Guan, W. Hu., VisComposer: A Visual Programmable Composition Environment for Information Visualization, Visual Informatics, Volume 2, 2018, pp. 71-81.
[9] S. Zhu, G. Sun, Q. Jiang, M. Zha, R. Liang, A survey on automatic infographics and visualization recommendations, Visual Informatics, Volume 4, Issue 3, September 2020, pp. 24-40.
[10] S. R. Midway, Principles of Effective Data Visualization, Patterns, Volume 1, Issue 9, 11 December 2020, 100141.
[11] O. Tymchenko, N. Kunanets, O. Khamula, S. Vasiuta, O. Sosnovska, Synthesis and research of a model of factors of infographics compositional design with elements of visual communication, Proceedings of the 2nd International Workshop on Intelligent Information Technologies & Systems of Information Security with CEUR-WS Khmelnytskyi, Ukraine, March 24–26, 2021, pp. 303-322.
[12] O. Tymchenko, S. Vasiuta, O. Khamula, A. Konyukhov, O. Sosnovska, M. Dudzik, Synthesis of Factors Model of Data Visualization in the Infographics, 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications Science and Technology PIC S&T'2019, Volume 2. (October 8-11). Kyiv, 2019. pp. 451-454.
[13] O. Tymchenko, O. Khamula, S. Vasiuta, O. Sosnovska, O. Mlynko, A Comparison of Methods for Identifying the Priority Hierarchy of Influencing Factors, Proceedings of the 3rd International Workshop on Intelligent Information Technologies & Systems of Information Security, Khmelnytskyi, Ukraine, March 23-25, 2022, CEUR Workshop Proceedings 3156, CEUR-WS.org 2022, pp. 228-237.
[14] V. Repeta, O. Havrylyshyn, I. Myklushka, Modelling of the Factors Influence on the Quality of the Digitization Process of Old Books by Fuzzy Logic, Proceedings of the 16th IEEE International Conference on Computer Science and Information Technologies, CSIT 2021, Lviv, Ukraine, September 22-25, 2021, Volume 2, pp. 182-185, Code 175940.
[15] V. Repeta, V. Zhydetskyy, O. Cherednichenko, I. Liakh, Modeling with fuzzy logic the migration capacity of UV-ink components as a factor of potentially harmful pollution of packaging, CEUR Workshop Proceedingsthis link is disabled, 2021, 2853, pp. 83–89.
[16] V. Repeta, Y. Kukura, V Kukura, Analysis of the relationship of quality factors in the solventless lamination process, Journal of Print and Media Technology Research, Volume 7, Issue 1, 2018, pp. 27 – 34.