Optimization of the Structure of an Information Security Textbook using Genetic Algorithm

Valerii Lakhno[1], Yurij Tikhonov[2], Nazym Sabitova[3], Lidiya Taimuratova[4], and Mereke Zhumadulova[4]

[1] National University of Life and Environmental Sciences of Ukraine, 6a Heroyiv Oborony str., Kyiv, 03041, Ukraine
[2] Luhansk Taras Shevchenko National University, 3 Koval str., Poltava, 36003, Ukraine
[3] L. N. Gumilyov Eurasian National University, 2 Satpayev str., Astana, 010000, Republic of Kazakhstan
[4] sh. Yesenov Caspian University of technology and engineering, 32 Microdistrict, Aktau, 130000, Republic of Kazakhstan

Abstract

The work shows that using a Genetic Algorithm (GA), it is possible to solve the following problems: eliminating parallel edges, cycles, loops, duplicating vertices with similar parameters, and other features of the structure of a Computer Ontology (CO) graph. Such CO optimization with the help of GA, for example, of the Information Security (IS) Subject Area (SBA), helps to eliminate logical contradictions that violate the integrity of CO and reduce the efficiency of the functioning and application of the Electronic Textbook (ET). It is also shown that the optimization of the IS SBA CO content is carried out to increase its information richness and ensure adaptation to the information needs of users through a periodic reduction in the volume of the SBA CO to specified limits. The solution was achieved by extracting IS CO elements whose semantic meaning is less than the required one. It has been established that the optimization problem for CO can be reduced to a discrete optimization problem, the algorithms for effective solutions of which using GA are known.

Keywords

Information security, electronic textbook, computer ontologies, optimization, genetic algorithm.

1. Introduction

In today's digital world, data plays a key role. Companies, governments, and even ordinary people store and transmit large amounts of information. Therefore, developing IS competencies will allow users to protect their data from leaks, hacker attacks, and other threats [1, 2].

Acquiring competencies in the field of information security has several key reasons and important aspects, among which are: the ability to recognize the signs of possible attacks and apply appropriate information security measures; compliance with regulations and laws, which will allow companies and organizations, while complying with information security rules, to avoid fines and negative consequences for their business processes; career development, since specialists with competencies in the field of information security can count on career growth [3, 4].

Note that the ontological approach in teaching information security can be very useful, starting from primary school age, since CO is a structured description of concepts, terms, and their relationships in a certain field of knowledge [5, 6]. The use of CO when forming the structure of, for example, an ET on

CPITS-2024: Cybersecurity Providing in Information and Telecommunication Systems, February 28, 2024, Kyiv, Ukraine
EMAIL: lva964@nubip.edu.ua (V. Lakhno); uruytihonov1@gmail.com (Y. Tikhonov); sab_nazym@mail.ru (N. Sabitova); taimuratova@mail.ru (L. Taimuratova); mereke.zhumadilova@yu.edu.kz (M. Zhumadilova)
ORCID: 0000-0001-9695-4543 (V. Lakhno); 0009-0003-9365-618X (Y. Tikhonov); 0009-0001-2531-5653 (N. Sabitova); 0000-0002-1692-4350 (L. Taimuratova); 0000-0002-2974-0227 (M. Zhumadilova)

information security issues can simplify the understanding of complex concepts related to information security and information security.

The use of CO when teaching the basics of information security allows to: structure the knowledge of students so that it is easier for them to navigate a large amount of information and understand the relationships between various concepts of information security; visualize the hierarchy of concepts and connections between them in the field of information security, which makes educational material more accessible and understandable, especially for schoolchildren; build various training cases and scenarios to better understand the relationships between vulnerabilities, threats and information security measures, etc.

All of the above served as the motivation for the study aimed at optimizing the structure of the ET for the IS course, designed for schoolchildren.

2. Procedures for Optimizing Computer Technologies of Electronic Textbooks on Information Security

The complex structure of relationships between the concepts reflected in the ontology of information security management systems, as well as its dynamic content during operation, requires the use of certain optimization procedures. These procedures are implemented to minimize response times to requests; meet the requirements, but not exceeding the space allocated for placing the ontology in computer memory; resolve conflicts between data entered from different sources, as well as meet other requirements and criteria that are to be determined during the development of specific information security controls.

We should not forget about the specific requirements dictated by the characteristics of, for example, a specific category of students. CO optimization is also carried out to adapt its content to the information needs of users, excluding elements that are rarely used or not used at all. For example, for schoolchildren, several concepts do not have to be taught as part of the information security program. For example, such complex categories as network protocols are not immediately perceived by older categories of students.

In terms of graph theory, CO structural optimization (elimination of conflicts, preservation of integrity, compliance with restrictions on the maximum volume) consists of alternating procedures for adding and reducing the ontology graph. This optimization affects parameters within a given range of values of the number of vertices in the case of maximizing the sum of the importance coefficients of its vertices and CO edges.

Checking the connectivity of a graph can be done using the consequences of the theorem on estimating the number of edges through the number of vertices and the number of connected components [7].

If we denote by p and q—the number of vertices and edges of the graph, respectively, the following two conditions must be met [7, 8]:

1. *if* $q > (p-1)(p-2)/2$ *then* the graph is connected.
2. for the connected graph it is true $-1 \le q \le p(p-1)/2$.

When making changes—adding new elements to the CO, CO modifying, or removing CO elements—the system must check the CO integrity. That is, the CO is checking for the absence of duplicate and/or mutually negative statements. This procedure can be implemented through a mechanism for detecting test reviews that are opposite in content by comparing them (contrasting them) in the case of sequential logical inversion of one of the review statements using the resolution method. If direct and inverted statements coincide, the system will receive a signal that integrity has been violated [5, 6].

Let's consider an example of one of the topics in the discipline "Information Security," and more specifically consider the topic—"Database Protection (DBP)." The example is taken as closely related to related topics of ET in information security, as well as to the disciplines: basics of programming, basics of algorithmization, etc.

Let's consider, for example, a CO fragment of the topic—DBP, see Fig. 1.

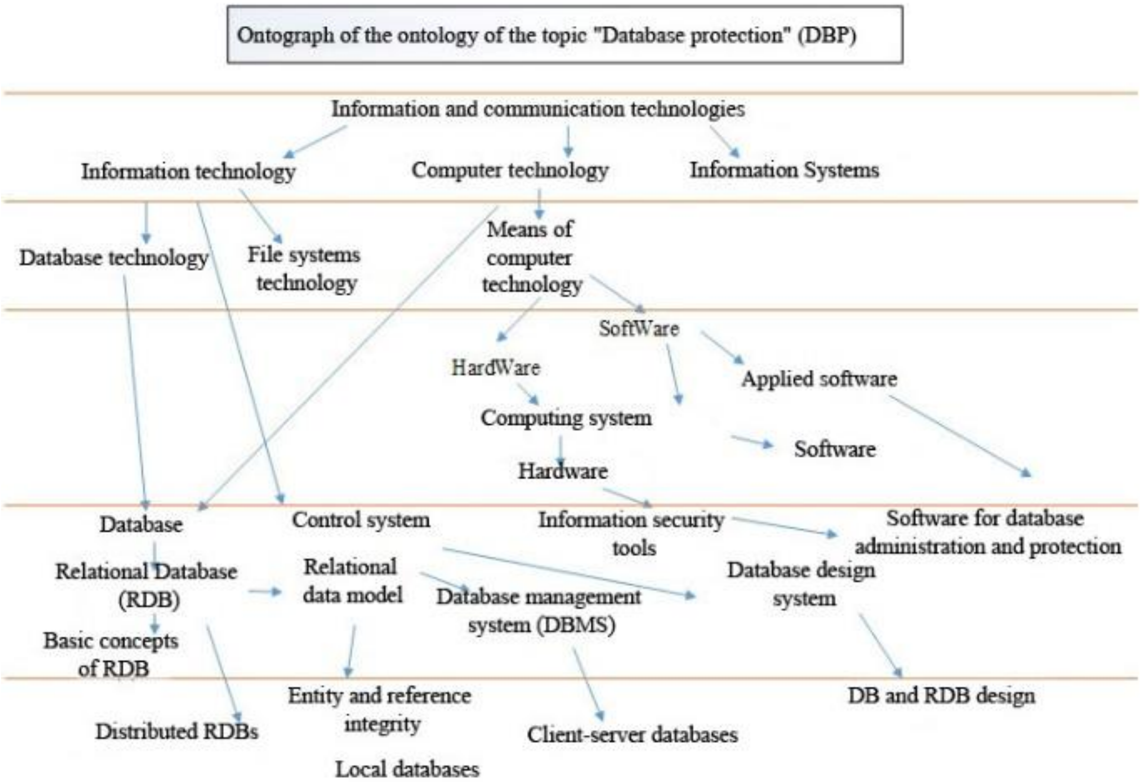

Figure 1: Fragment of the IS SBA CO ontograph (database protection topic)

In Fig. 1, the root concept of the ontograph is divided into levels of depth of knowledge about the SBA (horizontal orange lines). Moreover: the depth, heights of the concept (tier), and width of the graph are quite large. This is not a tree, so it is impossible to talk about the balance of the tree. Note that the presence of cycles also interferes with the perception of the ontograph. There is branching of the graph and edge density, which characterizes the proximity of the graph to a fully connected graph. Note that the variety in the number of connections in Fig. 1 is irregular. There is graph entanglement, which includes analysis of the ratio of the number of vertices with multiple inheritance to the number of all vertices in the graph and analysis of the average number of parent vertices for a vertex.

In general, if you include all the concepts of the fragment and their connections in the ET IS, then the ET will be relevant to modern knowledge about SBA, but difficult to perceive and weakly cognitive.

To improve perception, the expert (in this case, a teacher of the IS discipline preparing the corresponding ET) outlines, based on the syllabus and his experience, the level of depth of studying SBA, indicating on the ontograph the concepts up to which the student should study SBA—IS.

All higher-level concepts from those indicated up to the root are automatically included in the ET based on the CO description. To improve perception and for ET correct operation, the ontograph is divided into branches. Each branch leads from the root concept to the final concept specified by the expert. Graph processing methods are used.

The procedure for selecting the final concepts of a branch is poorly formalized. For example, it would be possible to set the depth of the SBA study (and the final concepts of the branches) according to the levels of the ontograph.

The educational SBA CO for the topic “Database Protection” includes about 200 concepts. It is irrational to work with a matrix of dimension 200, much less carry out manual procedures using an optimization algorithm. Therefore, for illustration, let’s take a small fragment of the ontograph. Let us select from the fragment (see Fig. 1) the vertices selected by the expert (see Fig. 2), and illustrate with them the matrix description and then the steps of the GA.

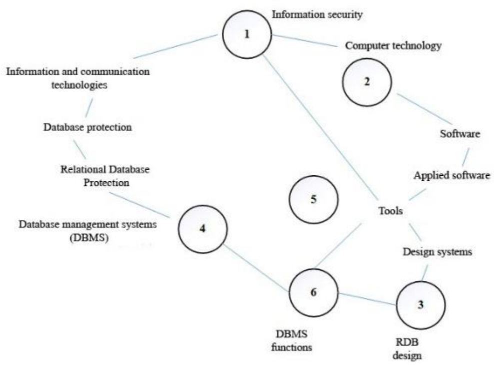

Figure 2: Expert-selected vertices (subjectively selected for example)

In Fig. 2, the selected vertices are numbered. Note that the selection of vertices was carried out subjectively to demonstrate the general concept of the possibilities of computer optimization of an ontograph using GA [9, 10].

Let's simplify the graph by removing the names of the vertices and leaving only their numbering. As a result, we get the following image, see Fig. 3.

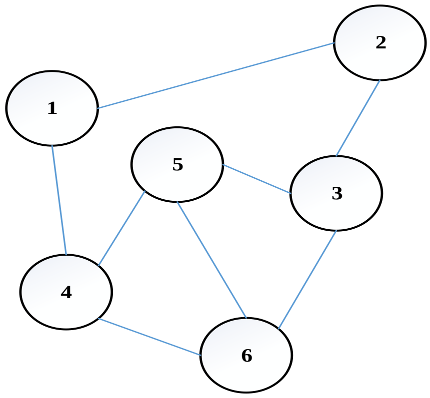

Figure 3: Initial graph for the GA

Based on the adjacency matrix, we encode the chromosome consisting of the vertices of the original graph. Based on a GA implemented in Python, see Fig. 4, a graph structure was obtained with a procedure for dividing the vertices of the graph of ET computer ontologies according to the information security course, see Fig. 5.

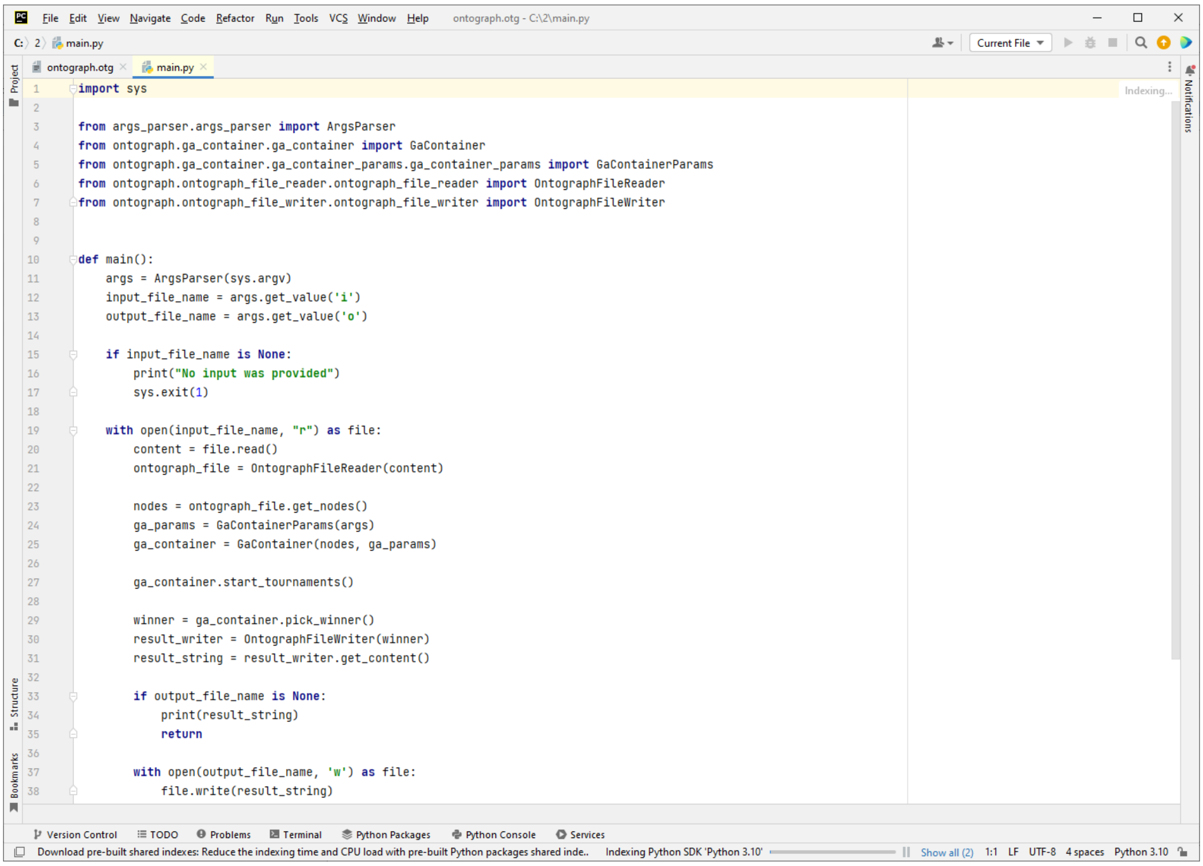

Figure 4: Example of implementation of a genetic algorithm for optimizing computer ontologies for ET on information security

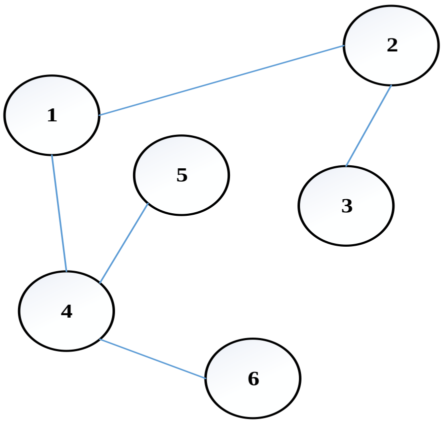

Figure 5: Diagram of the graph with the procedure for dividing the vertices of the CO graph for the topic under consideration

The proposed methodology for modifying CO for ET using the example of one of the topics of the IS SBA, by representing CO in the form of a graph and using techniques for working with the graph and finding optimal paths on the graph, in particular, GA can be used to automate the preparation of ETs in other areas. However, at this stage, there are no appropriate computer tools. In particular, to transform the CO OWL description into a computer description of the graph corresponding to the CO ontograph. Such a transformation manually is not possible due to the large dimension of the problem.

Therefore, for the full implementation of this methodology, it is necessary to create a tool that automates this procedure for ET development. This is a prospect for further research on this topic.

3. Conclusion

As a result of the study, the following results were obtained:

- It is shown that with the help of a GA, it is possible to solve the following problems: eliminating parallel edges, cycles, loops, duplication of vertices with similar parameters, and other features of the structure of a CO graph. Such CO optimization with the help of GA, for example, of the IS subject area SBA, helps eliminate logical contradictions that violate the CO integrity and reduce the efficiency of the ET.
- It is shown that the optimization of the IS SBA CO content is carried out to increase its information richness and ensure adaptation to the information needs of the user through a periodic reduction in the volume of the SBA CO to specified limits. The solution is achieved by extracting CO elements whose semantic meaning is less than the required one.
- It has been established that the optimization problem for CO can be reduced to a discrete optimization problem, the algorithms for effective solutions of which using GA are known.

References

[1] H. Hulak, et al., Dynamic Model of Guarantee Capacity and Cyber Security Management in the Critical Automated System, in: 2nd International Conference on Conflict Management in Global Information Networks, vol. 3530 (2023) 102–111.

[2] V. Grechaninov, et al., Formation of Dependability and Cyber Protection Model in Information Systems of Situational Center, in: Workshop on Emerging Technology Trends on the Smart Industry and the Internet of Things, vol. 3149 (2022) 107–117.

[3] P. Anakhov, et al., Increasing the Functional Network Stability in the Depression Zone of the Hydroelectric Power Station Reservoir, in: Workshop on Emerging Technology Trends on the Smart Industry and the Internet of Things, vol. 3149 (2022) 169–176.

[4] P. Anakhov, et al., Protecting Objects of Critical Information Infrastructure from Wartime Cyber Attacks by Decentralizing the Telecommunications Network, in: Workshop on Cybersecurity Providing in Information and Telecommunication Systems, vol. 3550 (2023) 240–245.

[5] M. Uschold, M. Gruninger, Ontologies: Principles, Methods and Applications, Knowl. Eng. Rev. 11(2) (1996) 93–136. doi: 10.1017/S0269888900007797.

[6] C. Nuntawong, C.S. Namahoot, M. Brückner, A Semantic Similarity Assessment Tool for Computer Science Subjects Using Extended Wu & Palmer's Algorithm and Ontology, Information

Science and Applications, LNNS 339 (2015) 989–996. doi: 10.1007/978-3-662-46578-3_118.
[7] A. Plahin, S. Kryvyi, N. Petrenko, Ontological methods and means of processing subject knowledge: monograph, Publishing House of the Volodymyr Dahl EUNU (2012).
[8] M. Lytras (2011). Software Technologies in Knowledge Society, Journal of Universal Computer Science, Vol. 17, no. P. 1219–1221.
[9] G. Alexandru-Lucian, A. Iftene, Using a Genetic Algorithm for Optimizing the Similarity Aggregation Step in the Process of Ontology Alignment, 9th RoEduNet IEEE International Conference (2010) 118–122.
[10] J. Wang, Z. Ding, C. Jiang, GAOM: Genetic Algorithm Based Ontology Matching, IEEE Asia-Pacific Conference on Services Computing (2006) 617–620. doi: 10.1109/APSCC.2006.59.