PRESENTATION OF CORPORATE KNOWLEDGE DURING THE TRANSITION TO HIGHLY AUTOMATED SYSTEMS

Ekaterina Mashina

Germany

ABSTRACT

This article addresses the pressing issue of developing methods for transferring corporate knowledge during the transition to highly automated technologies based on the use of effective knowledge representation models. The paper examines the practical implementation of existing approaches to model-based knowledge representation. As a result of the analysis, the author concludes that semantic models are the most effective way to represent corporate knowledge when preparing for the transition to highly automated technologies. Since the key process of semantic representation of corporate knowledge is the creation of subject ontologies, the author proposes schemes for implementing such processes, oriented towards different types of companies, and provides a description of the procedures carried out at various stages of these processes. The materials of the work are applicable for the development of practical methods for representing corporate knowledge, including when transitioning to highly automated systems.

KEYWORDS

Knowledge Transfer, Knowledge Representation, Semantic Knowledge Model, Domain Ontologies, Patterns

1. INTRODUCTION

In the context of the widespread transition to Industry 4.0, one of the key tasks is the effective preservation and management of corporate knowledge [1]. However, flaws in the synchronization of knowledge management systems that arise during the transition to highly automated technologies can lead to a loss of competitive advantage and a decline in the innovative potential of companies. Research shows that such mistakes can cause a decrease in the productivity of newly implemented technologies, which in turn leads to a 20-25% reduction in planned profits [2], while ineffective knowledge transfer procedures lead to repeated costs for training and adapting employees, reaching up to 30% of the total training budget. In addition, the loss of critical information that occurs when knowledge is incorrectly transferred to highly automated systems can hinder further innovation processes, potentially reducing the company's market share [3]. Thus, improper management of knowledge transfer in the context of the transition to Industry 4.0 technologies can result in losses for companies and pose a serious threat to their sustainable development [4].

The main gaps in the development of knowledge transfer mechanisms in highly automated systems are the insufficient integration of interdisciplinary descriptions of subject areas, the limited adaptability of existing models to dynamic conditions, and the lack of effective methods for verifying and validating knowledge. These limitations prevent the full potential of automated systems from being realized and require work to be done to create universal knowledge transfer mechanisms. In this regard, there is currently a high demand for work on

DOI: 10.5121/ijnlc.2025.14502

creating a methodology for describing corporate knowledge that ensures the smoothest possible transition to highly automated technologies.

A lot of research has been done on creating mechanisms for describing knowledge in automated systems. One of the first works in this area is by I. Nonaka and H. Takeuchi, who described the process of knowledge creation in organizations [1]. T. Davenport and L. Prusak also made a significant contribution by researching knowledge management and its impact on organizational effectiveness [2]. Other authors, such as M. Alavi and D. Leidner, have explored the impact of technology on knowledge management [5]. Important work has also been done by C. Choo, who studied internal knowledge transfer processes in organizations [4]. In addition, K. Sveiby researched methods for evaluating intangible assets and their connection to knowledge management [6].

The studies conducted emphasize the importance of creating effective interdisciplinary mechanisms for transferring knowledge to automated systems. At the same time, researchers unanimously agree that one of the most significant factors in organizing the process of effectively transferring knowledge to highly automated systems is the procedure for creating a knowledge representation model.

A properly created knowledge representation model allows corporate information to be structured, categorized, and organized correctly. This contributes to more accurate and faster retrieval of the necessary data, as well as the effective use of automated procedures. In addition, adequate models help reduce uncertainty and increase the reliability of decision-making in a dynamic environment, as well as allow the integration of diverse sources of information, which is a key aspect for achieving system compatibility. The use of correct knowledge representation models also leads to the optimization of data processing and improved interaction between system components.

At the same time, it should be borne in mind that, by the time of the transition to Industry 4.0 technologies, many corporate information systems had not yet developed sustainable approaches to corporate knowledge modelling. In this regard, before starting work on the transfer of knowledge to highly automated systems, it is necessary to develop the most appropriate approach based on an analysis of existing solutions for creating knowledge description models.

The purpose of the work described in the article is to conduct a comparative analysis of possible models of knowledge representation, identify their positive and negative qualities, assess their applicability for creating procedures for transferring corporate knowledge to highly automated systems, and formulate proposals for creating the most reasonable methods for such representation. The results of this study will be of interest to companies planning to transition to highly automated Industry 4.0 technologies, as well as organizations seeking to increase their competitiveness by improving the efficiency of corporate knowledge utilization.

2. METHODOLOGY

The choice of knowledge representation model in corporate information systems is determined by the need to ensure uniformity and simplicity of knowledge representation. Uniformity of representation facilitates the integration of different data sources and reduces the likelihood of errors [7]. The ease of understanding knowledge provided by intuitive models contributes to more effective interaction between users and the system and increases the “level of acceptance” of information technology in the company [2].

Several models of knowledge representation have been developed: logical model, frame model, production model, and semantic model.

The logical model of knowledge representation is an abstraction that describes the structure and relationships in terms of logical expressions and relations. The formal description of knowledge underlying the logical model can be represented as:

$$\varphi = \langle B, S, A, P \rangle$$

where B is a set of basic elements that make up the alphabet of knowledge representation, S is a set of syntactic rules that allow correct relationships to be generated and adequate expressions to be constructed on the basis of basic elements, A is a set of adequately generated statements, and P is a set of semantic rules that allow new adequate statements to be generated.

One of the main advantages of the logical representation model when used for knowledge transfer purposes is its ability to provide a high degree of independence from physical implementation. This simplifies the process of changing the structure of the knowledge being described [8]. Among the weaknesses is the difficulty of interpreting logical expressions for users who do not have specialized knowledge. This can limit the accessibility of information both during the transfer period and during further use [9]. In addition, a logical model may be less effective in processing large amounts of knowledge compared to physical models, which can negatively affect the performance of information systems [10]. Nevertheless, in some cases, the use of logical models remains relevant due to their ability to maintain the integrity and consistency of knowledge [7].

A frame model of knowledge description is a structure based on the concept of representing objects as a set of attributes and their values, which serve to organize information about objects, events, and situations as hierarchically related objects. The basis of this representation is a frame, which is usually understood as the following structure:

$$f\,[\langle v_1, g_1\rangle, \langle v_2, g_2\rangle, \ldots \langle v_k, g_k\rangle]$$

where f is the name of the frame, $\langle vi, gi\rangle$ is the i-th element of the frame, v_i is the name of the i-th element of the frame, and g_i is its value. Thus, a frame is a complete structure that, when all its elements are filled in, describes a fact. This allows for the modelling of extremely complex knowledge structures [11]. A significant advantage of representing knowledge in the form of frames is the ability to include assumptions and expectations in the description. This is achieved by assigning default descriptions of a whole set of different situations to frame slots. Another strength of the frame model of knowledge description is its ability to encapsulate knowledge and ensure contextual coherence. This greatly simplifies the process of information processing and generating conclusions based on it [12]. Among the disadvantages of frame representation is its limitation in representing dynamic changes and complex relationships between the objects being described. Another serious disadvantage of the frame model of knowledge representation is that frames can be prone to problems with generalization and specificity. This seriously complicates the modelling of more abstract concepts [13] and limits the universality of representation [14].

The production model of knowledge representation is based on the use of rules [15]. This model describes knowledge using so-called production rules, which are essentially programs consisting of a single operator of the form

IF <*condition*>, THEN <*action*>

This model allows knowledge to be described and processed effectively, providing flexibility in conclusions and adaptability [16]. Among the strengths of the production model are its modularity and ease of updating rules. This facilitates the integration of new knowledge [17]. The disadvantage of this model is potential performance issues with many rules and the complexity of managing conflicts between them [18]. In addition, production systems can suffer from insufficient expressiveness in describing complex interactions between objects and their states [14].

The semantic model of knowledge representation describes a subject area as a set of objects and the relationships between them. When using this approach, a specific subject area is described by creating a semantic network, which is a graph in which a set of vertices corresponds to the objects being described (concepts, events, processes), and a set of arcs corresponds to the relationships between objects.

This representation of knowledge allows information to be effectively structured and interpreted [19]. One of the main advantages of the semantic model of knowledge representation is its ability to express complex relationships and hierarchies. This contributes to a better understanding of the information environment [20]. A significant disadvantage of semantic knowledge representation models is the high computational complexity when processing large amounts of data [21]. In addition, semantic models require careful development and formalization of ontologies, which can lead to significant time and resource costs [22].

3. RESULTS

Today, semantic knowledge representation models have become the dominant paradigm in highly automated industrial systems due to their ability to effectively integrate heterogeneous data. These models provide a high level of abstraction, allowing systems to extract meaning from data rather than simply storing it in a structured form [12]. In addition, semantic models improve interaction between different system components by providing a common language for describing knowledge. Another important aspect is the ability to automatically interpret and infer new knowledge, which is critical for creating real-time systems [23]. Modern semantic technologies, such as Resource Description Framework (RDF) and Web Ontology Language (OWL), allow the creation of flexible and extensible models, which makes them particularly attractive for dynamically changing production environments. Thus, with the widespread transition to Industry 4.0 technologies, semantic models are a powerful tool for improving the efficiency and adaptability of highly automated systems [22]. Ontologies are the main method of description and structuring when using semantic representation. Creating adequate ontologies for subject areas is a key task in the use of semantic models [24]. Building ontologies is a complex and time-consuming task. Therefore, numerous studies are being conducted to find common ways to improve the efficiency of this process, as well as on the creation of a variety of software tools for performing specific tasks. It should be noted that the type, representation, subject area, and other parameters of corporate information can vary significantly even for closely related companies. Therefore, effective ontologies for describing corporate knowledge can also differ significantly from one another. In this regard, the construction of each corporate ontology is largely a unique task that each company must solve. A generalized diagram of the automated process for solving such a task is a sequential execution of six consecutive steps (see Figure 1).

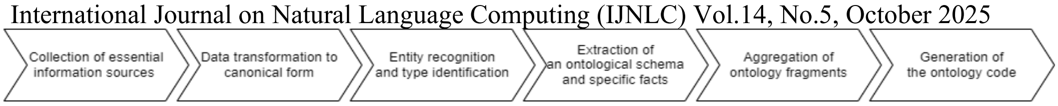

Figure 1. Key stages in the formation of ontologies based on the analysis and transformation of corporate information

The stage of collecting relevant information sources is one of the most important parts of the process of creating a corporate knowledge ontology, the success of which determines the subsequent quality and applicability of the solution being created. The task of this stage is to select documentary sources that most fully describe the functioning of the company.

Most companies today have and actively use centralized data repositories, including those organized as ECM (Enterprise Content Management) systems. These solutions contain a significant number of documents related to the company's operations in their centralized repositories. However, a significant portion of corporate information is also distributed in specialized enterprise information systems that contain a significant portion of corporate knowledge and are not always connected to the company's centralized repository [25].

The task of the data transformation stage to a canonical form is to organize the transition from documents with an arbitrary structure to a uniform text representation. This transformation includes the following phases: recognition, functional and structural analysis. This stage involves the removal of various "junk" (i.e., non-letter and non-number) character values, the removal of multiple spaces and text markup characters, the decryption of acronyms, the identification of units of measurement, and similar tasks. The task of the entity recognition and type determination stage is to perform the procedures of extracting and recognizing named entities contained in the texts processed at the previous stage. Currently, a significant number of productive tools have been developed to perform such actions. One of them is Stanford Named Entity Recognizer (Stanford NER), which is a powerful tool for automatic recognition of named entities in texts. It is based on machine learning methods and allows models to be trained on specific data in various fields. Stanford NER stands out for its high accuracy and flexibility in customization for specific tasks, as well as the availability of pre-trained models for different languages. Analogues to Stanford NER include SpaCy, OpenNLP, and NLTK.

The task of the ontology and fact extraction stage is to obtain ontological fragments in the form of a set of classes, their relationships, object properties, and datatype properties, as well as specific instances (facts) that describe a specific subject area based on the information obtained in the previous step. The generation of the desired ontology fragments is carried out in two sub-steps. First, an ontological schema is extracted from a set of texts reduced to canonical form. In this case, hierarchical relationships between headings are interpreted as object properties representing relationships between the classes being formed. In the second sub-stage, specific instances of named entities are extracted from the texts. The result of this sub-stage is fragments of an ontological model, which must then be aggregated, including by performing operations to refine the names, properties, and relationships of classes, as well as procedures for their possible separation or merging.

The task of the actions performed at the stage of aggregating ontology fragments is to integrate the obtained fragments into a unified ontological model designed to represent and store knowledge extracted from various sources. As part of these actions, classes with identical names are merged and consolidated, followed by the formation of a common set of values for object properties and instances. Duplicate classes with identical property structures and names are removed. At the same time, it is proposed to carry out the procedure of merging classes with similar names using a method based on the Levenshtein distance, which determines the degree of similarity between two class names [26]. If classes with the same name are found, new object properties are created.

During the code generation stage, machine-readable ontology code is created based on the ontological model obtained (for example, in OWL2 DL format). After final "manual verification," the OWL code of the ontological model representing corporate knowledge generated in this way can be supplemented or modified using ontological modelling editors.
Approach to creating an ontological description of corporate knowledge described above is based on existing universal tools. Therefore, it can be argued that companies wishing to reduce the risks of knowledge loss when transitioning to highly automated systems can use the proposed sequence of actions to do the work themselves.

4. ANALYSIS OF RESULTS

It should be noted that the process of gradually creating ontologies based on primary sources provides a good quality description of objects and their relationships within the subject area under consideration, but this approach is quite labour-intensive. Therefore, creating a description "from scratch" in this way is only justified for ontologies describing small, rapidly developing areas of knowledge that are characterized by many innovative objects and relationships. For large companies with many technological processes, work centres, and their interrelationships, the process of step-by-step ontology creation based on primary sources may prove to be excessively labour-intensive [27]

If it is possible to reuse typical description fragments, it seems reasonable to rely on the technology of constructing the ontology being created using primary blocks (patterns) [28]. The workflow for carrying out work according to this scenario can be presented in the form shown in Figure 2.

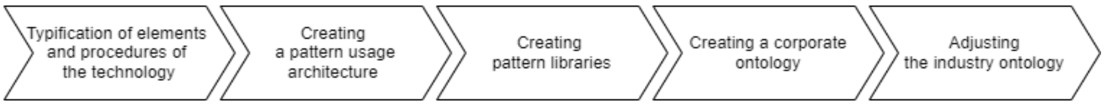

Figure 2. The main stages of ontology formation based on the application of patterns

At the pattern typification stage, elements and procedures for using typified fragments are categorized and structured to create an ontological model describing the process of designing technology for the subject area. When developing unified procedures for such technology, one can rely on the classification of objects used, proposed in the NeOn project. Based on this, the ontology of ontological design patterns can be used when implementing the technology (see Figure 3).

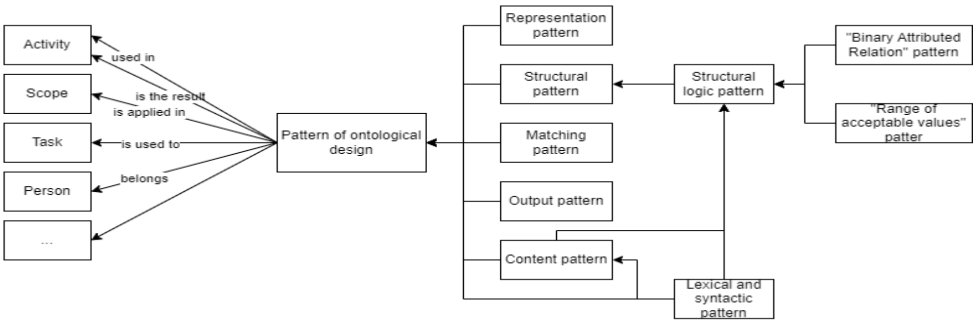

Figure 3. Basic ontology design patterns

When creating pattern usage technology, consideration should be given not only to the possibility of reusing previously created patterns for describing corporate processes and objects (based on the use of pattern libraries), but also to the processes of creating new patterns and modifying them to consider the variability of their application context (see Figure 4).

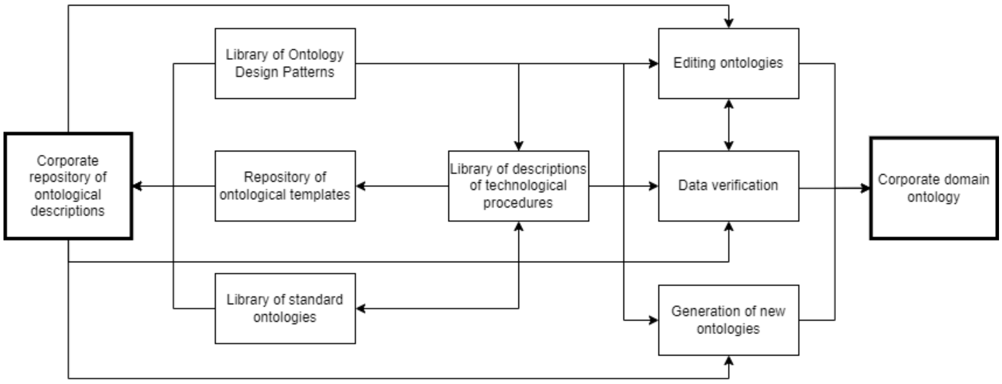

Figure 4. Technological scheme for creating ontologies based on the use of unified patterns

To obtain high-quality results in ontological design when creating pattern libraries, all the features of the production and technological operations carried out at the company's work centres, as well as all the interrelationships between them, must be considered as much as possible. To do this, the scheme for building ontology "from scratch" shown in Figure 1 can be used. At the same time, to minimize the labour costs of creating unique ontology fragments, it is necessary to "atomize" the description of production processes as much as possible. This will allow to identify the most frequently repeated fragments of descriptions that are suitable for formatting as standardized patterns suitable for multiple applications. At the same time, the process of creating a unified corporate subject ontology can be represented as the integration of separate ontologies of elementary subject areas, created based on a unified set of pattern libraries. In this case, the resulting ontology will be a unidirectional multi-connected graph. This will allow the creation of subject area hierarchies with multiple inheritance (see Figure 5), the application of which can be particularly effective in cases of multidisciplinary corporate knowledge.

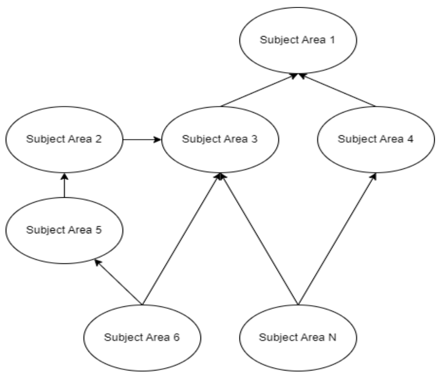

Figure 5. Example of integrating subject areas into a unified representation

The proposed approach to creating a corporate ontology in the form of a node-based assembly of individual patterns, pre-developed with the characteristics of the enterprise's technological processes in mind, will significantly accelerate the process of describing corporate knowledge to a level suitable for implementing knowledge transfer procedures in highly automated technologies. This will significantly reduce the risks of losing critical corporate knowledge during the transition to Industry 4.0 technologies, as well as preserve the volume of knowledge that is not in demand in the updated technologies as potential for further technological innovations.

The creation and further use of a structured representation of corporate knowledge, both for the specific task of transferring company data to highly automated systems and for supporting the entire business of the company, provides significant competitive advantages. However, it is more appropriate to work on the creation of ontological representations of knowledge using unified pattern technology within the framework of the concept of "open knowledge engineering," which involves the creation of publicly available technological patterns and elementary technologies in subject areas, combining which it is possible to move on to the creation of both industry-specific and cross-industry technologies that can significantly facilitate the processes of widespread use of innovative knowledge.

5. Conclusions

The article discusses approaches to creating a knowledge representation model designed to ensure corporate knowledge preservation procedures during the transition to highly automated Industry 4.0 technologies. It proposes an approach to building corporate knowledge models based on the application of subject ontologies, which are constructed using universal patterns. The following conclusions can be drawn from the study, which are relevant for the further development of practical methods and techniques for preserving and effectively using knowledge during the transition to Industry 4.0 technologies:

- to create models for describing corporate knowledge intended to organize the transition to highly automated technologies, the most reasonable approach is the semantic representation of knowledge based on ontological representation,
- it is advisable to create complex ontologies describing multidisciplinary corporate knowledge by focusing on an approach based on the use of universal patterns, which are the most unified descriptions of ontological fragments of corporate knowledge,
- it is advisable to conduct the process of creating corporate ontological representations of knowledge in an open information space within the industry, which allows the expert competencies of a large number of specialists to be combined.

References

[1] I. Nonaka and H. Takeuchi, "The knowledge-creating company: How Japanese companies create the dynamics of innovation," Long Range Planning, vol. 29, no. 4, p. 592, Aug. 1996. [Online]. Available: http://dx.doi.org/10.1016/0024-6301(96)81509-3

[2] T. Davenport and L. Prusak, Working Knowledge: How Organizations Manage What They Know, 01 1998, vol. 1.

[3] G. Krogh and S. Grand, Justification in Knowledge Creation: Dominant Logic in Management Discourses, 01 2000, pp. 13–35.

[4] C. Choo, "The knowing organization: How organizations use information to construct meaning, create knowledge and make decisions," International Journal of Information Management, vol. 16, no. 5, p. 329–340, Oct. 1996. [Online]. Available: http: //dx.doi.org/10.1016/0268-4012(96)00020-5

[5] M. Alavi and D. E. Leidner, "Review: Knowledge management and knowledge management systems: conceptual foundations and research issues," MIS Q., vol. 25, no. 1, p. 107–136, Mar. 2001. [Online]. Available: https://doi.org/10.2307/3250961

[6] K. Sveiby, The New Organizational Wealth: Managing & Measuring Knowledge-based Assets, ser. Berrett-koehler Series. Berrett-Koehler Publishers, 1997.

[7] C. Batini and M. Scannapieco, Data and Information Quality: Dimensions, Principles and Techniques, ser. Data-Centric Systems and Applications. Springer International Publishing, 2016.

[8] C. J. Date, An introduction to database systems. Pearson Education India, 2006.

[9] M. Jarke, M. Lenzerini, Y. Vassiliou, and P. Vassiliadis, Fundamentals of Data Warehouses. Springer Berlin Heidelberg, 2003. [Online]. Available: http://dx.doi.org/10. 1007/978-3-662-05153-5

[10] R. Elmasri, Fundamentals of database systems. Pearson Education India, 2008.

[11] M. Minsky, A Framework For Representing Knowledge. Berlin, Boston: De Gruyter, 1979, pp. 1–25. [Online]. Available: https://doi.org/10.1515/9783110858778-003

[12] R. Brachman and H. Levesque, Knowledge Representation and Reasoning, 05 2004.

[13] R. Schank, Dynamic Memory Revisited. Cambridge University Press, 1999.

[14] R. J. Brachman, "What is-a is and isn't: An analysis of taxonomic links in semantic networks," Computer, vol. 16, pp. 30–36, 1983.

[15] A. Newell, Human Problem Solving. USA: Prentice-Hall, Inc., 1972.

[16] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 3rd ed. USA: Prentice Hall Press, 2009.

[17] J. Laird, A. Newell, and P. Rosenbloom, "Soar: An architecture for general intelligence," Artificial Intelligence, vol. 33, pp. 1–, 09 1987.

[18] K. D. Forbus, "Qualitative process theory," Artif. Intell., vol. 24, pp. 85–168, 1984.

[19] R. J. Brachman and J. G. Schmolze, "An overview of the kl-one knowledge representation system," Cognitive Science, vol. 9, no. 2, pp. 171–216, 1985. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0364021385800148

[20] N. Guarino, Formal Ontology in Information Systems: Proceedings of the 1st International Conference June 6-8, 1998, Trento, Italy, 1st ed. NLD: IOS Press, 1998.

[21] C. Bishop, Pattern Recognition and Machine Learning, 01 2006, vol. 16, pp. 140–155.

[22] N. Noy and D. Mcguinness, "Ontology development 101: A guide to creating your first ontology," Knowledge Systems Laboratory, vol. 32, 01 2001.

[23] R. Kowalski, Logic for problem-solving. NLD: North-Holland Publishing Co., 1986.

[24] E. Mashina, "Identification and formalization of the implicit personal knowledge of the employee acquired by him in the course of work," in 14th International Conference on Natural Language Processing and Computational Linguistics (NLPCL 2025), September 27 - 28, 2025, Toronto, Canada, vol. 15, no. 17, 2025, pp. 13–24.

[25] E. Mashina and B. Pavel, "The influence of descriptive language practices on the process of external integration of corporate knowledge," Lecture Notes in Networks and Systems, vol. 636, pp. 104–119, 2023.

[26] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals." Soviet Physics Doklady, vol. 10, no. 8, pp. 707–710, 02 1966, doklady Akademii Nauk SSSR, V163 No4 845-848 1965.

[27] A. Sattar, E. Salwana, M. Surin, M. Ahmad, U. Malaysia, R. Khan, M. Ahmad, and A. K. Mahmood, "Comparative analysis of methodologies for domain ontology development: A systematic review," International Journal of Advanced Computer Science and Applications, vol. 11, 01 2020.

[28] A. Gangemi and V. Presutti, Ontology Design Patterns, 05 2009, pp. 221–243.

AUTHOR

Ekaterina Mashina, Master of Science degree in 2024, with a thesis on the use of artificial intelligence to formalize tacit knowledge. Main research interests include semantic analysis, cognitive linguistics, and neurotechnology. Author of more than 15 publications in peer-reviewed scientific journals, participant of international conferences. In 2022, was awarded the title of "Best graduate of ITMO University in the field of science".