Use of the Smart City Ontology for Relevant Information Retrieval

Kunanets Nataliya[1] [0000-0003-3007-2462], Halyna Matsiuk[2] [0000-0002-8857-1857]

[1] Lviv Polytechnic National University, Lviv, Ukraine
[2] Ternopil Ivan Puluj National Technical University, Ternopil, Ukraine
nek.lviv@gmail.com
galuna.matsiuk@gmail.com

Abstract. Today, many Intelligent Web Applications require the use of distributed knowledge bases formalized in the form of ontologies and thesauri. They are the new intellectual means for searching resources on the Internet, new methods for representing, systematizing and processing of knowledge and inquiries. They can accurately and effectively describe the data semantics for certain subject area and solve the problem of incompatibility and contradictory concepts. A brief overview of recent foreign investigations aimed at solving the problems of ontologies construction and development in various fields of science is carried out in this paper. Different points of view on the concepts of ontology used in modern information technologies are covered and the definition of this term is given in this paper. The ontological approach to the formation of information inquiries is considered, the search scheme on the basis of ontologies, the algorithm for the ontology domain construction and the example of ontology fragment construction in the "Smart City" data domain are presented.

Keywords: ontology, data domain, thesaurus.

1 Introduction

The effectiveness of scientific-research works depends directly on the quality of their information support. The development of modern information technologies is characterized by steady increase in the information amount and tasks related to their application in various fields. This results in difficulty of finding the necessary information in mass data of the global information spacewhich are needed to solve specific problems. Hence there is a problem of obtaining relevant information which meets the user needs. This is also affected by the fact that modern retrieval systems seek information without taking into account the words semantics on demand, as well as the context in which they are used.

In recent years, it has been suggested to build and use systems which use ontologies for relevant information retrieval in order to seek necessary knowledge, information objects or information resources containing information necessary for users to solve specific problems. This makes it possible to solve the problems related to the

information retrieval in the open Web environment faster and making it dynamic, of high-quality and user-friendly.

The formation of new socio-economic structures i.e. city-states (sociopolicies) on the basis of smart decisions is taken place in the world recently. The "Smart city" is one of the concepts of such development. It is the model of the city development based on the modern information and communication technology use. With the increase of the information flow there is the need of finding new ways of its storing, representing, formalizing and systematizing as well as its automatic processing. The use of knowledge databases representing the model or concept of knowledge storage is the most efficient. Comprehensive knowledge databases contain not only real-world information, but also the inference rules, resolve automatic conclusions concerning new facts and as a result meaningful information processing. This results in the growth of interest in information resources one of which is the ontology, a new step in knowledge and data presentation and integration in artificial systems [1]

Data domain ontology is used to clarify and define its general terminology base. It is an "agreement on the sharing of concepts (terms) including means of domain knowledge presentation and agreement on logical input methods. It can be regarded as the formal description of the worldview in a particular area of interest consisting of a set of terms and rules for the use of these terms limiting their meaning within the specific data domain"[2].

The advantage of the ontology as the way of representing knowledge is its formal structure, which makes the computer information processing easier.

The objective of the paper is to investigate the use of ontological projects by foreign researchers in various fields and to present the methodology of the "Smart city" data domain ontology.

2 Recent research and publications analysis

There are many investigations aimed at solving the problems of the construction and development of ontologies in various fields of science.

In the paper [3] the problem of automatic learning in ontology engineering is explored, because there is still a lack of fully automatic approach from a text corpus or dataset of various topics to form ontology using machine learning techniques. In this paper, two topic modeling algorithms are explored, namely LSI & SVD and Mr.LDA for learning topic ontology. The objective is to determine the statistical relationship between document and terms to build a topic ontology and ontology graph with minimum human intervention. Experimental analysis on building a topic ontology and semantic retrieving corresponding topic ontology for the user’s query demonstrating the effectiveness of the proposed approach.

We have a number of research in the field of medicine.

In the paper [4] an object of the research is the cardiovascular system because it is "one of the most committed to the development of diseases associated with modern life". Authors describe a histological ontology of the human cardiovascular system developed in collaboration among histology experts and computer scientists. The

histological ontology is developed following an existing methodology using Conceptual Models and validated using OOPS! (OntOlogy Pitfall Scanner!): An On-line Tool for Ontology Evaluation, how accurately the ontology can answer the Competency Questions. It is publicly available at http://bioportal.bioontology.org/ontologies/HO and https://w3id.org/def/System. The histological ontology is developed to support complex tasks, such as supporting teaching activities, medical practices, and bio-medical research or having natural language interactions.

In the paper [5], the authors developed an upper-level ontology SCTO and to use it as the basis for defining the terms in Systematized Nomenclature of Medicine-Clinical Terms – a comprehensive medical terminology used for standardizing the storage, retrieval, and exchange of electronic health data. The SCTO is implemented in Web Ontology Language (OWL) 2, to support automatic inference and consistency checking. The approach will allow integration of SCT data with data annotated using Open Biomedical Ontologies (OBO) Foundry ontologies, since the use of OGMS will ensure consistency with the Basic Formal Ontology, which is the top-level ontology of the OBO Foundry. Currently, the SCTO contains 304 classes, 28 properties, 2400 axioms, and 1555 annotations. It is publicly available through the bioportal at http://bioportal.bioontology.org/ontologies/SCTO/. The resulting ontology can enhance the semantics of clinical decision support systems and semantic interoperability among distributed electronic health records. In addition, the populated ontology can be used for the automation of mobile health applications.

In the works of foreign scientists we notice a number of investigations in the field of education, especially electronic one.

In the paper [6] the ontology for syllabus representation is presented. The paper [7] presents an ontological model called PLET4Thesis, which has been designed in order to organise the process of thesis development using the elements required to create a personal learning environment.

In [8, 9] authors explore the use of ontology for knowledge representation in knowledge-based recommender systems for e-learning

Many works aim to explore ontology-based approaches for smart city mobility application.

The Intelligent Transportation System is a part of the mobility aspect of a smart city which necessitates the demands for seamless interoperability with information processing needs in focus.

Intelligent transportation systems are used to improve the performance and safety of road transportation of a smart city. A crucial element for the success of this system is the exchange of information, not only between vehicles, but also among other components in the road infrastructure through different applications. One of the most important information sources in this kind of systems is sensors. In the papers [10, 11] an ontology-driven a four layer architecture to improve the driving environment through a traffic sensor network is proposed. The system performs different tasks automatically to increase driver safety and comfort using the information provided by the sensors.

In the works [12, 13] authors propose to use the SSN ontology to manage the sensor information in an intelligent transportation architecture. The system was tested in a traffic light settings application, allowing to predict and avoid traffic accidents, and also for the routing optimization.

In the work [14] we have an approach for an ontology-based Smart Bike Sharing System (SBSS) architecture that addresses the challenge of unauthorized and illegal parking. The heterogeneous data collected from different sensors at the respective stations are transferred to the central server and integrated using the ontology.

3 Ontology as knowledge storage systems

The term ontology was used for the first time in philosophy and was associated with Aristotel, who tried to classify objects for knowledge system development. There are many definitions of this concept. The most widely used in philosophy is the following definition: ontology (from the Latin ontologia comes from the ancient Greek onto - existing, something that exists, and the Greek logoc - the doctrine, science) - this is the doctrine of being, the branch of philosophy, which clarifies the fundamental problems of existence, the development of entity, the most important. In other words, ontology is a science of being, a science about the nature of things and the interconnection between them [15].

Despite the fact that the term "ontology" comes from philosophy, in computer science it has an independent meaning. The main differences are in the fact that the ontology in computer science has a formalized representation language which computer can process without human involvement; and an ontology is created for solving the certain problem, and therefore it is evaluated rather from the application point of view than from its completeness.

One of the most well-known ontology definitions was introduced by Thomas R. Gruber: "An ontology is an explicit specification of a conceptualization" [16].

Another important ontology definition is the definition offered by Guarino, based on the analysis of seven different definitions: "Ontology is a logical theory which gives an explicit, partial account of a conceptualization" [17].

Under "conceptualization" we mean the description of the system of concepts, objects and relationships connecting them with each other. It divides any area of knowledge existing in integral form, separates from this area individual objects, and then formulates the relations particular for this area. The main part of the formally presented knowledge is based on conceptualization.

T. Gruber and N. Guarino ontology presentation provided the basis for ontology definitions of the data domain by other researchers according to their specific area of professional interests. Eduard Hovy represents the following definition as the working ontology definition: ontology is data structure with the given in it characters enabling to represent conceptualization for processing by computer programs [18].

A. Palagin, S. Kryviy, N. Petrenko define ontologies as formal representation of conceptual knowledge about the data domain and in its significance is relevant to the intellectual information system knowledge base and its architecture represents specif-

ic form of human thinking. It (thinking) uses judgments, statements, concepts and relations between them in the cognitive development. The latter is the foundation based on the construction of the component of the scientific theory part i.e. ontological knowledge base in a certain problem area [19].

Ontologies can be used as the agent between the user and the information system, they make it possible to formalize the terminology arrangement. Ontologies are used as data sources for many computer programs (for information retrieval, text analysis, knowledge extraction and other information technologies) providing more efficient processing of complex and diverse information.

The main purpose of ontologies is the information integration. Ontologies connect two important aspects: first, they define the formal information semantics providing processing of this information by computer, and, secondly, determine the real world semantics enabling on the basis of general terminology to link the information presented in the form necessary for computer processing, with information presented in the form convenient for human perception [20].

Ontologies are new intellectual means for resources search on the Internet, new methods for representing and processing knowledge and inquries. They can accurately and effectively describe the data semantics for a certain data domain and solve the problem of concepts incompatibility and inconsistency. Ontologies have their own processing tools (logical output), relative tasks of semantic information processing. Hence, due to ontologies while accessing the search engine, the user has the opportunity to receive responses of resources semantically relevant to the inquiry (Figure 1).

The ontology of the data domain is as a rule the data structure providing the possibility of storing data domain knowledge by interpreting the terms of the data domain and interpreting such types of paradigmatic relations as a "part-whole", "class-subclass", and some types of associative relationships [21].

In general, the ontology structure is a set of elements of four categories: the concept (classes, concepts); samples; axioms; relations (relationships), attributes (properties).

4 Ontological presentation of the "Smart City" data domain knowledge

The ontology construction usually starts with the dictionary of terms used to describe a certain area of knowledge, creation of the system for accurate definition of these terms, as well as documenting of the basic logical relationships between the terms. Thus, the ontology includes a set of terms and rules, according to which these terms can be combined for representation of knowledge in a certain area.

Let us consider the stages of ontology construction for the "Smart city" data domain presentation.

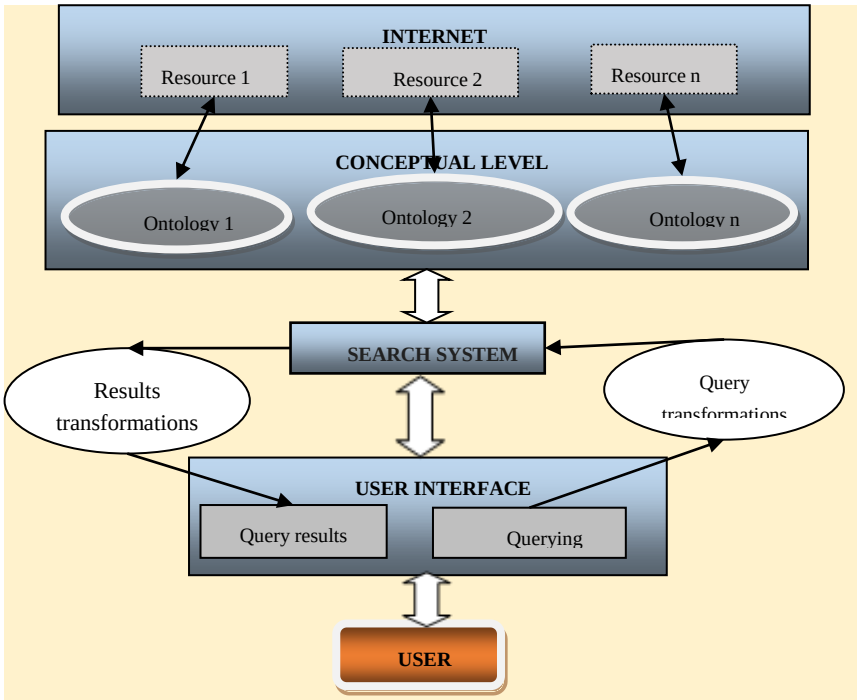

Fig. 1. The retrieval pattern based on ontology.

According to IDEF5 (Integrated Definition for Ontology Description Capture Method) approach designed for ontological analysis of the data domain, analysis of the basic terms and concepts used to describe the objects and processes of a certain data domain, limits of their use and interconnections between them, the process of ontology construction consists of five main stages:

1. Organize and Define Project is the stage involving the goal defining and context of the ontology development project as well as distribution of roles between the project members;
2. Collect Data is the stage where collection and accumulation of initial data necessary for the ontology construction are carried out;
3. Analyze Data is the stage focused on analyzing and grouping of collected data for the terminology construction;
4. Develop Initial Ontology is the stage where preliminary ontology is formed on the basis of selected data;
5. Refine and Validate Ontology is the activity including ontology refinement and verification, it completes the development process [22].

The methods of data domain ontology construction include three stages:

1. Preliminary analysis of the data domain.
2. Construction of the ontology categorial level.
3. Graphical design of the data domain ontology.

System analysis of the data domain covers the following tasks: formation of systematic representation of knowledge about the given data domain; selection of the most significant, relevant concepts of the data domain; forming the data domain glossary. According to IDEF5 standard the main concepts of the data domain ontology investigations are used:

- while constructing the conceptual model, domain knowledge in the form of the set of concepts and relationships connecting them are used, each concept has its name and can have attributes, each attribute can have its meaning taking into account the data domain peculiarities.
- the model can be constructed using the vocabulary of terms used in describing the characteristics of objects and processes of the corresponding data domain, precise and unambiguous definitions of all the terms of this dictionary and the classification of logical interrelations between these terms;
- the basis of the design of the information system conceptual model is the data domain with perfect and verified terminology system, well-formed hierarchy of definitions of subject-oriented concepts [22].

The main tasks to be solved by ontologies are:

- sharing of the general structure understanding by people and software agents;
- accumulation and reuse of knowledge in the data domain;
- the possibility of making explicit assumptions in the data domain;
- separation of the data domain knowledge from operational ones;
- analysis of the data domain knowledge [23].

The cognitive-information component in the formulation and solution of the "Smart city" data domain tasks is currently based on the development of intelligent systems for the analysis of native language texts. The main element of the semantic analysis system is the effective vocabulary database. One of the tasks of creating intellectual systems of the data domain ontology is a set of concepts and relationships between them [24].

The authors of the investigation formed the "Smart city" bilingual thesaurus for presentation and exchange of knowledge in this data domain.

Complex-structured "Smart city" data domain is the part of real world, consisting of the group of objects reflecting a set of concrete and abstract concepts with certain connections between them. Like any data domain, it is characterized by the following properties: the availability of sections described in different, but similar systems; hierarchical structure of subsections.

The terms list construction is the most important moment in the construction of the data domain ontology. The "Smart city" thesaurus terms were used as the basis of the elements set.

The relation set (connections) is also based on the results of the preliminary "Smart city" domain analysis. It is necessary to construct a set of arcs connecting the directed ontograph top points. The ontograph top points are the set of the data domain concepts. The ontograph top (without categorial ontology level) is the generic term which has no superclass, and at the lower level there are specific concepts without specific notions in the given data domain. It is important to note that there is no single correct hierarchy of classes.

Let us give an example of the construction of the "Smart City" domain ontology fragment. The basic concept here is the "Smart city". Let us consider its definition: the smart city is an effective integration of physical, digital and human systems in artificial environment in order to provide stable, prosperous future for its citizens. The Smart City model is the system of classifications where the "Smart cities" can be developed and evaluated due to six different features: "Smart Mobility", "Smart Economy", "Smart Environment", "Smart People", "Smart Life", "Smart Government". Hence we distinguish the main classes: "Smart Mobility", "Smart Economy", "Smart Environment", "Smart People", "Smart Life", "Smart Government".

Using the process of top-down development for the class hierarchy, the "Smart city" ontology classification tree with one root vertex with the "Smart city" concept as denotation and taxonomic relation "genus – species" is built. The following relation is typical for the "Smart City" concept: the "Smart city" will exist providing the availability of the following "smart" categories: mobility, economy, environment, people, life and government.

Thus the taxonomy for the classification of the investigated data domain sections at the stage of the first level representation is obtained (Fig. 2).

As the result of further instantiation of each class, the creation of semantic networks for each term combined with each other due to hierarchical and associative relationships into a single network including the domain concept is carried out.

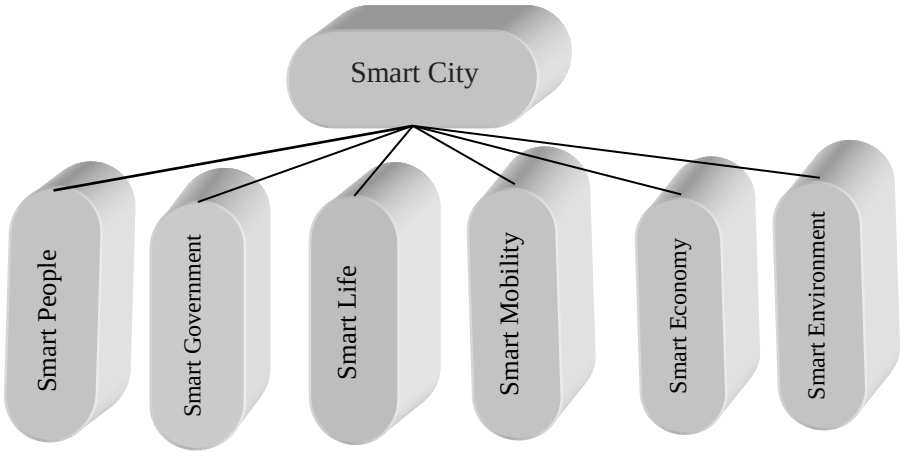

Fig. 2. Taxonomy of the "Smart city" data domain sections classification at the stage of the first level presentation

Ontogram for the "Smart city"data domain fragment consisting of the following concepts is shown in Fig. 3.: 1 – "Quality of Life", 2 – "Health care", 2a – "Telemedicine", 3 – "Smart Education", 4 – "Safety", 5 – "Smart House", 6 – "Low-energy hous", 7 – "Zero carbon building", 8 – "Zero energy building", 9 – "Energy-plus house", 10 – "Passive House", 11 - "Active House", 12 – "Alternative Material", 13 – "Ventilation", 14 – "Sensor", 15 – "Gas Sensor", 16 – "Heat Sensor" , 17 – "Smoke detector", 18 – "Light sensor", 19 – "Flame sensor", 20 – "Sensor of broken glass", 21 – "Geothermal heating", 22 – "Smart system", 23 – "Building management systems", 24 – "Energy Management System", 25 – "Central heating system", 26 – "Smart Windows", 27 – "Thermal insulation", 28 – "Eco-building", 29 – "Green roofs".

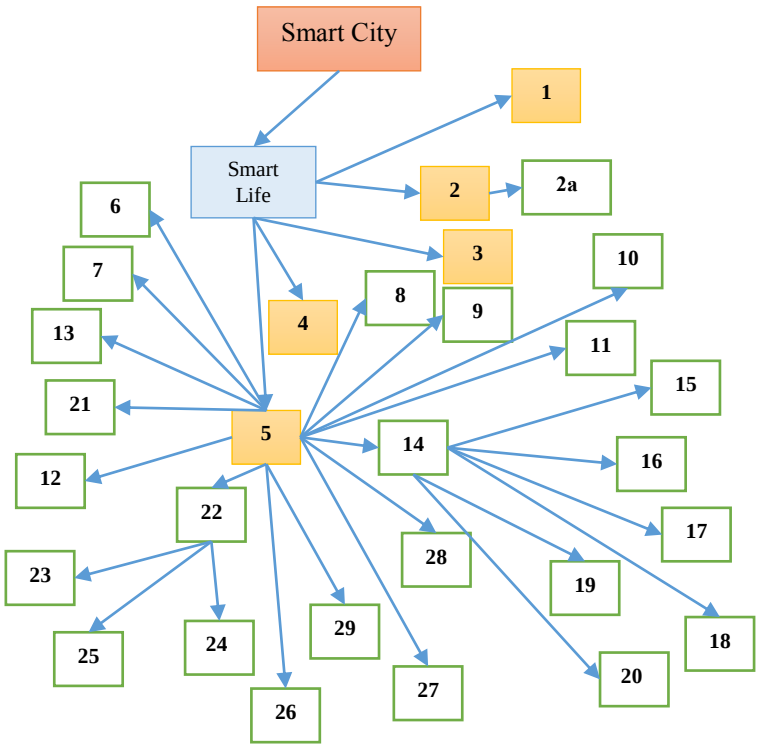

Fig. 3. Ontogram for the "Smart city"data domain fragment

The relation set of the offered ontograph consists of the elements: {class-subclass, whole- part, influences, consists of, uses, defines, possesses, forms, controls, etc.}

The construction of ontology as an interoperable model of knowledge representation in the process of scientific research of the "Smart City" data domain provides an appropriate formalized description of objects in the given area. Ontology defines the terms of this data domain, gives their interpretation, contains statements that restrict the meaning of these terms. The set of the data domain terms forms its terminology system represented as the "Smart city" thesaurus - a complete systematized set of data

about this domain knowledge, allowing its navigation by a person or computer. Ontology provides the thesaurus for presenting and sharing knowledge about a given data domain and numerous relationships established between the terms of this thesaurus.

The ontological approach used by us for presenting the concepts peculiar for the investigation field contains the basic set of terms. The formed thesaurus is used to describe the terms in the "Smart City" investigations. Ontology makes it possible to represent the basic concepts in such a way that they become basic for formal information retrieval, and the formal language of the ontologies description used for their coding, contributes to improving the efficiency of data processing.

5 Conclusion

The use of ontology is effective in retrieval and connection of information from different sources and environments. The "Smart city" ontology determines the terms by which we can describe and structure the data domain. The ontology language is used to provide accurately defined information and is the common set of terms for describing and presenting the given data domain.

The "Smart city" data domain ontology obtained as the result of modeling can be used as the basis for creation the knowledge base about the "Smart city". The model offers a systematic view of the city smart-concept and can assist in the evaluation how problems and initiatives affect city development.

The offered "Smart city" ontological model is aimed to conceptualize knowledge about this domain, to provide interrelated elements of the investigated data domain in a single system and to simplify the information retrieval.

The ontological model of the "Smart city" data domain in the context of information retrieval integrates the concept of various aspects of the "Smart city" architecture which can be used for communication by program agents and users and due to the built-in relationships between concepts simplifies the mechanism for necessary information retrieval relevant to the user request. The use of the ontological approach makes information retrieval dynamic, of high-quality and user-friendly.

References

1. Kunanets, N., Pasichnyk, V., Matsyuk, H., Matsyuk, O.: Ontological investigation of the "Smart city" data domain. In: 6st International Scientific-training Conference: Mathematics. Information technologies. Education, pp.48-52 (2017)
2. Hladun, A., Rohushyna, Yu.: Fundamentals of the thesaurus creation methodology using ontological and mereological analysis. Artificial Intelligence, 5, pp. 112-124 (2008)
3. Ran, M., Dhar, A., Vyas, O.: Semi-automatic terminology ontology learning based on topic modeling. Engineering Applications of Artificial Intelligence, 63, pp. 108-125 (2017)
4. Mazo, C., Salazar, L., Corcho, O., Trujillo, M., Alegre, E.: A histological ontology of the human cardiovascular system. Journal of Biomedical Semantics 8 (1), 47:1–47:15 (2017)

5. El-Sappagh, S., Franda, F., Ali, F., Kwak, K.: SNOMED CT standard ontology based on the ontology for general medical science. BMC Medical Informatics and Decision Making 18, 76 (2018)
6. Tapia-Leon, M., Rivera, A, C., Chicaiza, J., Luján-Mora, S.: The Use of Ontologies for Syllabus Representation. In: 17th International Conference on Information Technology Based Higher Education and Training (ITHET), pp.127-133. IEEE, Olhao Portugal (2018).
7. Tapia-Leon, M., Chicaiza, J., Santamaria, T., Luján-Mora, S.: Ontology for the personal learning environments in the development of thesis projects. In: 9th International Conference on Education Technology and Computers, pp.183-187 (2017)
8. Tarus, J. K., Niu, Z., Ghulam, M.: Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. Artificial Intelligence Review archive 50 (1), pp. 21-48 (2018)
9. Tarus, J. K., Niu, Z., Yousif, A.: A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining. Future Generation Computer Systems 72, pp. 37-48 (2017)
10. Susel, F., Hadfi, R., Takayuki, Ito, Marsa-Maestre, I., Juan, R.: Velasco Ontology-Based Architecture for Intelligent Transportation Systems Using a Traffic Sensor Network. Sensors 16(8):1287 (2016)
11. Susel, F., Takayuki, Ito: Ontology-Based Architecture to Improve Driving Performance Using Sensor Information for Intelligent Transportation Systems. Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems, pp. 245-255 (2018)
12. Susel, F., Takayuki, Ito: Semantic Integration of Sensor Data with SSN Ontology in a Multi-Agent Architecture for Intelligent Transportation Systems. IEICE Transactions on Information and Systems 12, pp. 2915-2922 (2017)
13. Susel, F., Hadfi, R., Takayuki, Ito: Architecture for Intelligent Transportation System Based in a General Traffic Ontology. A GPU-Based Pencil Beam Algorithm for Dose Calculations in Proton Radiation Therapy, pp. 43-55 (2016)
14. Patel, A. S., Ojha, M., Rani, M., Khare, A., Vyas, O. P., Vyas, R.: Ontology-Based Multi-agent Smart Bike Sharing System (SBSS). In: IEEE International Conference on Smart Computing (SMARTCOMP), pp. 417-422 (2018)
15. Lytvyn, V., Pasichnyk, V., Yatsyshyn, Yu.: Intelligent Systems: Novyy Svit-2000, Lviv (2008)
16. Gruber, T. R.: A translation approach to portable ontologies. Knowledge Acquisition, 5(2), pp. 199-220 (1993)
17. Guarino, N.: Formal Ontology and Information Systems. IOSPress, Amsterdam (1998)
18. Hovy, E.: Methods for Automated Ontology Building. Information Sciences. Ontologies: lecture. London (2004)
19. Palagin, A. V., Kryvyi, S. L., Petrenko, N. H.: Ontological Methods and Means of domain knowledge processing: Monograph, Luhansk (2012)
20. Kudryashova, E. E.: Methods and models of information systems design (monograph). In: International Journal of Applied and Fundamental Research, 6, pp. 117-118 (2010)
21. Berezhansyi, V. N., Potapova, E. E.: Classification of connections between notions in ontology in the physics of magnetic phenomena, MegaLing, pp. 12-21 (2010)

22. Menzel, C., Mayer, R.: IDEF 5 Ontology Description Capture Method: Concept Paper, Research Institute for Computing and Information Systems. Univ. of Houston-Clear Lake (1990)
23. Noy N.F., L. McGuinness, D.: Ontology Development 101: A Guide to Creating Your First Ontology'. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880 (2001)
24. Kunanets, N., Matsyuk, H.: Thesaurus of the subject domain "Smart City". Bulletin of "Lviv Polytechnics" National University, 872, pp. 179-189 (2017)