Semantic, Digitization, Design and Implementation of Ontology in Social Internet-Services

Nazish Mumtaz[1], Abida Begum[1], Bushra Gul[1], Salma Noor[1], Roman Odarchenko[2], Igor Machalin[2] and Olha Saliieva[3]

[1] Shaheed Benazir Bhutto Women University, Peshawar, Pakistan
[2] National Aviation University Kyiv, Ukraine
[3] Vinnytsia National Technical University, Vinnytsia, Ukraine
`odarchenko.r.s@ukr.net`

Abstract. Due to natural disasters, climatic changes, and lack of resources with the local heritage preservation authorities. Pakistan, in general, is losing some very important historical records. The heritage preservation is of immense importance to preserve our history and culture for the subsequent generations. With current advent in technology, the digitization of heritage has become cost-efficient and can be adopted by the relevant departments with the help of right expertise. This research uses the idea of Interoperable framework, machine readability and semantic representation for accessing digitally preserved information regarding the life history of Buddha from Gandhara Art work preserved at the museum of Peshawar. To enable semantic interoperability, we propose the Buddha domain ontology. It ensures the possibility of extracting meaningful and relevant data from a semantically represented repository of relevant data. The work opens ways to exploring Buddha life history and the data related to it in creative ways. This work will facilitate both the specialist and the novice web user the ease of access to desired information.

Keywords: Ontology, Semantic interoperability, Semantic web, Culture heritage, Buddha Ontology.

1   Introduction

Civilization is the initial period of human life style. Which is converted to generation to generation. The history of human life style and its related things provide the information. Specially that information which are the reason of change of human life. The initial history of Pakistan Indus valley civilization is also known as Harappa civilization. The seven historic periods were found from Mohenjo-Daro. One is the first period from 2500 to 3800 BC second are the middle period from 1700 to 2500 BC and the third are the last period from 1300 to 1700 BC was found from Mohenjo-Daro. Harappa was the big city of this civilization and the second was Mohenjo-Daro. The presence of these big cities reveals that these areas were very populated. The food was produced in large quantity because of fertile ground. This civilization reveals the presence of large forest, the shape of animals on the utensils and the bones show that

there were large number of animals which is used for agriculture and transportation. Harappa archeology show that there was a big factory of wheat grind which used to grind wheat. Big way of transportation of those people was boats, horses and camels. The fun and industrial life of the Indus valley civilization is pottery, instrument, and terracotta statue, marble statue [1].

The aim of this research is to preserve the history of Gandhara art. Pakistan need such kind of research, because of floods, earthquake, terrorism and similar hazards are eradicating our history so there is a need to preserve the historical rich culture of Pakistan. Due to modernization and less care from the local heritage preservation authorities, the culture history record of Pakistan, in general, is losing some very important historical record. This research will support the idea of protection of history. If the heritage preservation is not given its importance any sooner, the subsequent society will not even be able to explore it through archeology. With current advent of information technology, the digitization of any type of information has become very affordable and can be adopted by the relevant departments with ease. This will not only preserve our past and changing culture but also produce job opportunities. The basic outcomes of this research target the preservation of Gandhara art history in the best possible means, therefore we chosen 90 objects of Buddha artifacts from Gandhara art, to semantically digitized by implementing ontology. The ontology will overcome the issue and providing better access and knowledge discovery. This ontology will be able to express information among people and reuse of domain knowledge access and knowledge discovery [2].

1.1 Ontologies

"Ontology is the branch of "Meta Physics" concern with nature and relation of being". Ontology is decided what is of significance in an area and how information about it is arrange. Ontology is an explicit specification of conceptualization [3].

Ontology are used to describe the concepts in the specific domain and also describe relationship between these concepts.

1.2 Uses of Ontologies

Communication. Ontology arranges the framework in organization which abridged the theoretical confusion so the ontology gets better the communication among people [3].

Interoperability. Ontology in field provides the enterprise model and multi agent design for creation of incorporated toolkit [4].

Reliability. Ontology develop the reliability of system [3].

Reusability. Ontology should maintain the reusability thus   unlike model of system can be import or export [4].

The problem is to preserve the Pakistan history and culture by semantic digitization of Peshawar museum. Indus valley civilization is important and rich all over the world they were populated area because of fertilized ground and this valley was famous for trade and merchants came from different cities to exchange the merchandise and stayed. The problem in Pakistan is that Pakistan is not the main stream of global research so the project is initiative to place the Pakistan in global research. We choose ninety objects of Buddha artifacts for semantic digitization to fulfill this goal we im-

plement in this research an ontology which will enhance support for semantic digitization (converting analog data into machine readable form) of Peshawar museum. The ontology will overcome the issue and providing better access and knowledge discovery. This ontology will be able to express information among people and reuse of domain knowledge access and knowledge discovery.

2   Related Work

World Wide Web contains millions of documents which allow the people to retrieve the information from the database catalog. We need a search engine to extract the information. Now days there are many search engines to search the information but difficult to retrieve meaningful information to reduce this problem in search engine to extract meaningful information intelligently we use semantic web technology to provide better search engine.

2.1   Semantic Web

Michele R. Ramos [3], project is to design and validate the activities of life, it provides a generalized ontology of life events. The project used different previous approaches such as prosopography project. The traditional prosopography is related to social trend among culture and time period in order to understand the society. The prosopography project database is specific for custody, property, income, ethnic or social relationships. It cannot support the generalized context of life events. The biography light ontology supports the general biographic text provide open outline for linking biographic events with diverse resources of semantic web Biography vocabulary only distinguish four basic events birth, marriage, death and bio Event.

Semantic web is a collection of linked information in such a way which is easily process able by a machine on a global scale. "Semantic web is an extension of the current web" [3]. The goal of semantic web is machine process able an understandable information on the web. The purpose of semantic web is to share and reuse information on the web. This literature review introduces the semantic web technology to support the intelligent search engine [1], [2].
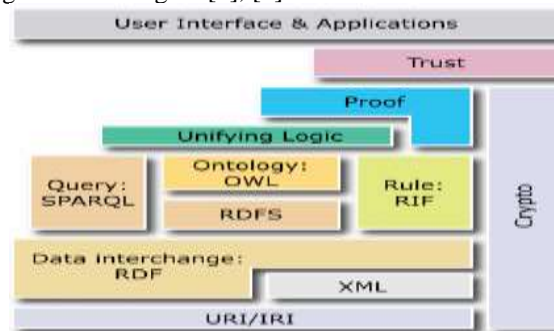

Fig. 1. Semantic Web Framework

Semantic web technology plays important role in semantic web that allow the meaning of data is to precise. Data on the semantic web is represent by using W3C standards RDF (Resource description frame work) and Owl (Ontology web Language) representation model which are used to represent ontology. This technology is used to support interoperability, automation and reuse of data. Current web contains large database but difficult to understand for the machine which is provided by user. When information is added on the web, we face research problem in search engine such as how search engine retrieves meaningful information intelligently. Semantic web solved this problem by semantic web technology to provide meaningful information by SPARQL query and domain ontology [4].

3 Motivation

The aim of this paper is to preserve the history in the best possible means using I.T. To achieve this, the available record will be some form of data, which will be made available with user friendly G.U.I. The prototype preservation system can be developed in following steps: Digitization of available heritage resources; Converting database of manual records into digital form; Discover new knowledge and digitize it with reference; Develop database and front end or API for online access; Explore the method to improve accessibility and usability.

Museum digitization motivate the API developer to create the API for Buddha Gallery of Peshawar museum, API developer will be able to create a meaningful application which present the objects in the form of narration and through this application the visitor will take interest in history and increase knowledge of visitor about history of subcontinent so in this way the civilization of Indus valley will be long lasting.

4 Proposed System Architecture

To reduced problems of existing system we have design and implement ontology on Buddha life story. We use semantic web technology to semantically digitizing these objects. For semantically digitizing we designed and implement Buddha ontology.
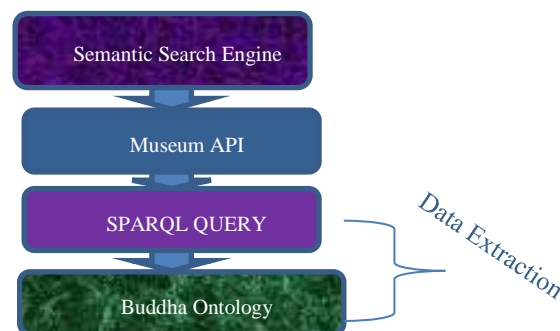

Fig. 2. Proposed System.

Buddha ontology contains thirteen concepts which describe the Gautama Buddha life events. We use own vocabularies to describe the relationship between instance of each class. We use the approach of arc2 triple store to mapped Buddha ontology into RDF format. Arc2 triple store provide SPARQL endpoint interface to extract data using SPARQL query engine. Proposed system shows the flow of data in a system. It shows the system work. Arc2 store load Protégé ontology for storing in the form of triples. MySQL extract the data from file maker pro and load in arc2 triple store here the information is store in RDF format. SPARQL query engine extract the data from arc2 triple store by using SPARQL query language. SPARQL query engine provide the user interface. See Fig. 3.
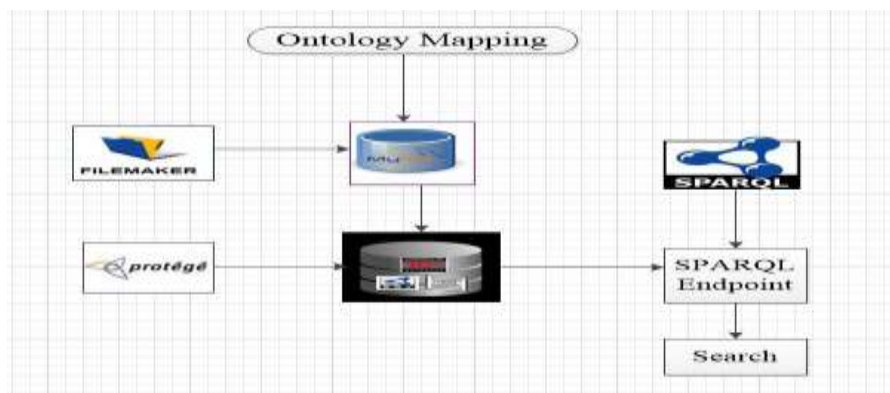

Fig. 3. Design Scenario of Buddha Ontology

4.1   Work Flow Development Process of Ontology

We developed an ontology of Gautama Buddha life story by the following process. The purpose of developing this ontology is to narrate the life events of Gautama Buddha or "Siddhartha". We use this ontology for application which we create in future work. The following steps were taken to implement Buddha ontology.

- Domain.
- Enumerated terms.
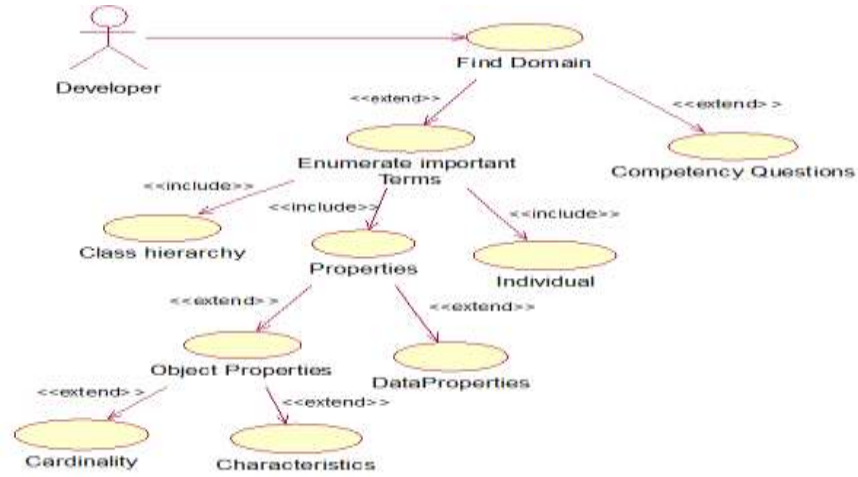- Class hierarchy.
- Properties
- Individual

Fig. 4. Ontology development work flow

5 Advantages of Proposed System

After semantic digitization these data available online on web which can be easily accessible and reuse. It facilitates the user to easily access data. It improves the search engine capability. It provides interoperability between one are more applications. User can use Buddha ontology for creating suitable application.

6 Implementation of Buddha Ontology

In this section we discuss the ontology development process, ontology mapping, relational database and Xampp sever installation steps. We need following parts for implementation.

Fig. 5, show the ontology mapping process using arc2 triple store. First of all, it takes data from file maker pro to create relational database using Mysql server. For storing this information in the form of triples arc2 triple store is takes Protege ontology and traditional database to store in RDF format. SPARQL endpoint queried the RDF data from arc2 triple store by using SPARQL query engine.
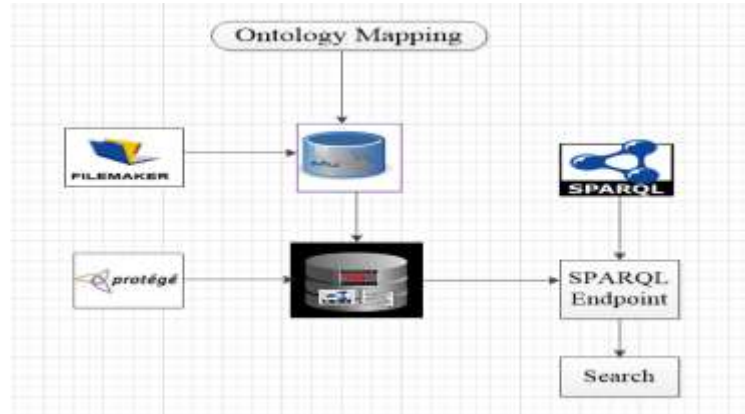
Fig. 5. Ontology Mapping Process

6.1 Ontologies Development

Problem Contextualization. The developed ontology is useful for more reason to share common structure information among people to reuse of domain information, to make domain supposition unambiguous, to examine area facts and to separate the domain knowledge from operational knowledge. In paper [5], the author using proté-gé for developing ontology. The artificial intelligent define many definitions of ontology. Define the concepts classes and property in area arranging the classes in hierarchic form. The ontology progress is the iterative process. There are many methodologies of developing ontology. Ontology is the model of designing reality of the world. Seven steps of developing ontology as follow as:

- Domain.
- Scope.
- Competency questions.
- Enumerated terms.
- Class hierarchy.
- Properties.
- Individual

These steps are important during ontology development process. The domain of ontology is describing what the ontology is covered. The scope of ontologies explains the use of ontology for an application. Competency question is the type of question which answer by developing ontology. Enumerated terms are important for selecting properties and classes. Class hierarchy are the set of classes which we select from terms to use in ontology Properties are also select from enumerated terms for using in ontology [6].

Dataset. For this project we have taken data from Peshawar museum in which we have chosen 90 objects of Gandhara art gallery to narrate life story of Gautama Buddha [10].

Shortcomings. Data which we collect for semantic digitization from Peshawar museum database is not complete because of more fields in database of every objects is missing. Metadata of each objects is not properly defined. Description of objects is too short. Objects are not properly labeled to overcome these problems we collect 90 objects for semantic digitization to describe on semantic web for reuse [11].

Mining of Plan Texts. Text Mining is the process of extracting information from the documents (PDF, XML, Plain text e.g.). We are use GATE software for text mining to extract information from the document "Buddha". Buddha document is plain text document, for knowledge discovery that we have created to appending all the information from the museum record about the Buddha gallery in which include Buddha artifacts that represent Gandhara art of stone.

6.2 Buddha Ontology

We plan to use Buddha ontology for the application which narrates the life story of Siddhartha. Basically, different concepts of Buddha life story are existing which is not meaningful. But over ontology generate the new meaningful concepts related to tinny objects. The aim is to cover different events of life of Buddha and define the concept of each term used in the story.

6.3 Class Hierarchy

We determine classes from important enumerated terms of Buddha life story which show in figure.

Fig. 6. Class Hierarchy of Buddha Ontology.
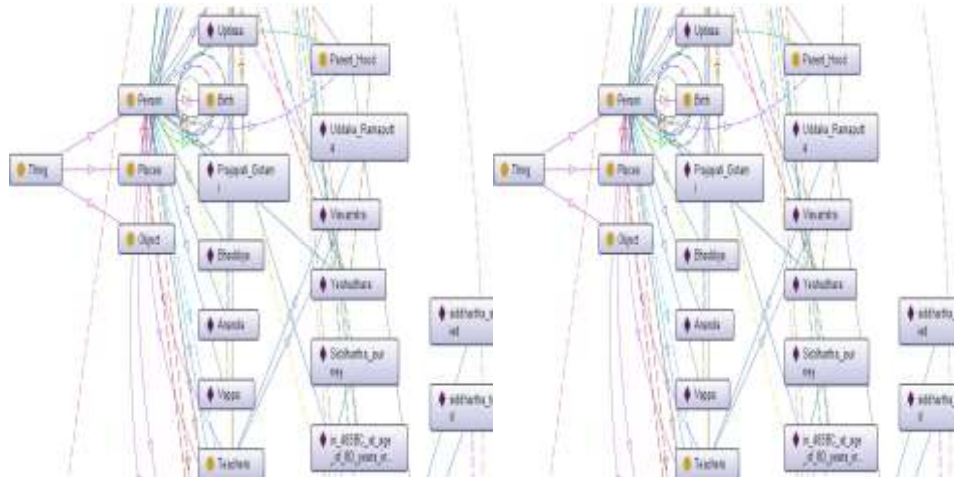
6.4   Ontology

Fig. 7. Buddha Ontology

6.5   Natural Language Process

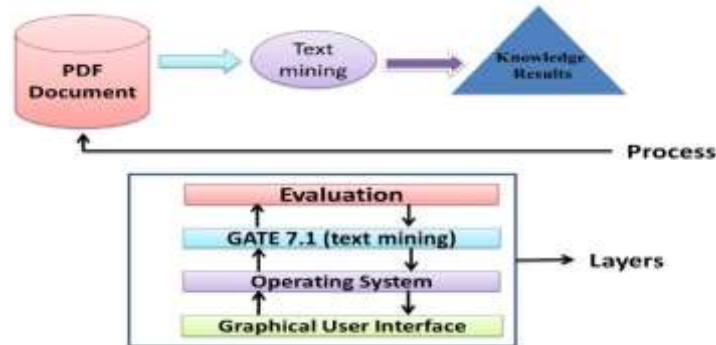The NLP process comprises of the following main steps as discussed in the fig below:

Fig. 8. NLP Process

NLP Steps on Buddha Document. Following are the Steps:
1. Load Buddha document in GATE.
2. Create corpus (corpora are a collection of documents, useful when processing multiple document) for Buddha document.
3. Select ANNIE (stand for A Nearly-New Information Extraction to extract information) and run application with corpus of Buddha.
4. Select processing resources (is a set of parameters).

5. Run pipeline with selected resources (pipeline is a set of selected parameters that set for document).
6. Create data stores (data store is used to store annotated documents).
7. Store Buddha document to data store.

Fig. 9. NLP on Buddha Documents

6.6   Results of NLP on Buddha Document

IE gate/extract facts and structured information from Buddha document collection. IE returns knowledge in the form of name entity. Name entity is much important to identify name in text and their classification into predefined categories of interest: Persons; Organizations (companies, government organizations, committees etc.); Locations (cities, countries, rivers etc.); Date and time expressions; Addresses; Sentences; Splitters; Various other types as appropriate.
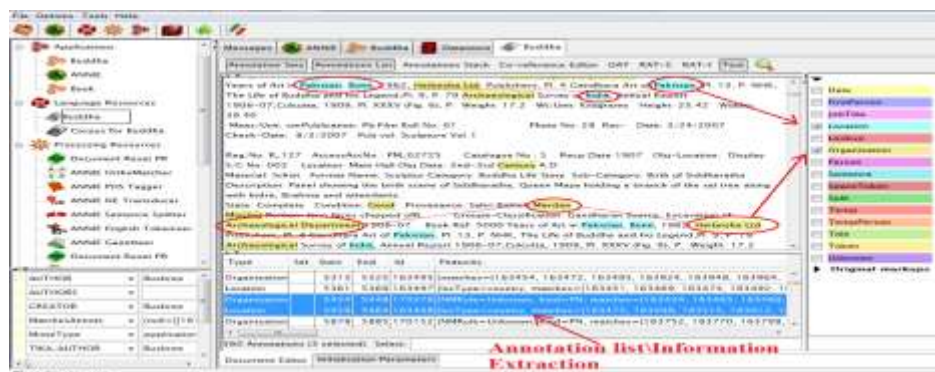

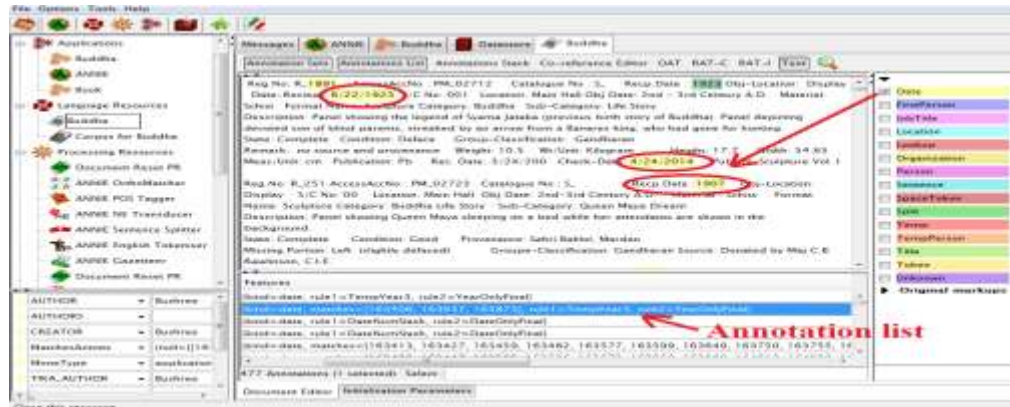**Fig. 10.** Information Extraction from Buddha Document

Fig. 11. Information Extraction from Buddha Document

7 Ontology Evaluation and Validation

Ontology evaluation and validation is important to check during development of ontology to verify that the ontology fulfills the requirement and check the incompleteness and redundancy of classes and instances.
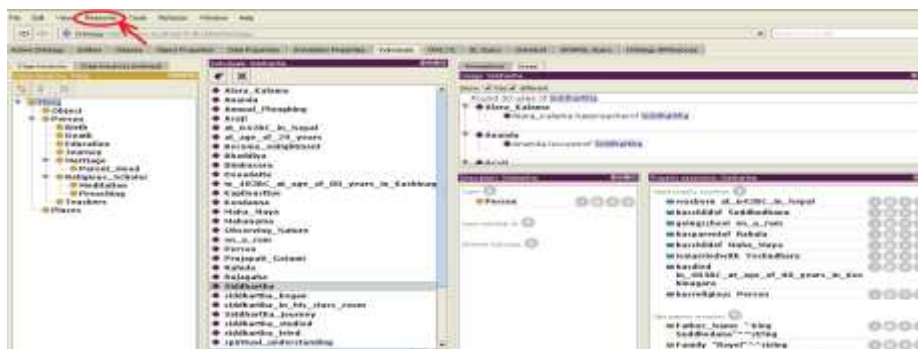

Fig. 12. Buddha Ontology Validation and Evaluation

8 Conclusion

This work will be useful for reusing and help support better exploration of the artifacts and new knowledge discovery and will eventually be used as a major contribute of the Peshawar museum API developers. Buddha ontology will make this history more explicit and will be help in developing API. Our future work will include API development. Ontology of Buddha life story main gallery of Peshawar museum which is useful for semantic digitization and also mapped the relational database into RDF data. The common information about Buddha life story is identified which is useful for the preservation of Gandhara art and Buddha ontology make these histories more

explicit help in developing the application which narrate the Buddha life in future work. Buddha ontology cover the events of whole life of Buddha. It is related to biographic light event ontology but it is specific for Buddha life events. Ontology pass through the GATE software which generate new concepts.

References

1. Jacob, Elin K. "Ontologies and the semantic web." Bulletin of the American Society for Information Science and Technology 29, no. 4 (2003): 19-22.
2. Xu, Xin, and Hubo Cai. "Semantic frame-based information extraction from utility regulatory documents to support compliance checking." In Advances in Informatics and Computing in Civil and Construction Engineering, pp. 223-230. Springer, Cham, 2019.
3. Khan, Shakeel Ahmad, and Rubina Bhatti. "Semantic Web and ontology-based applications for digital libraries: An investigation from LIS professionals in Pakistan." The Electronic Library 36, no. 5 (2018): 826-841.
4. Endara, Lorena, G. Burleigh, Laurel Cooper, Pankaj Jaiswal, and Marie-Angélique Laporte. "A natural language processing pipeline to extract phenotypic data from formal taxonomic descriptions with a focus on flagellate plants." (2018).
5. Noy, Natalya F., and Deborah L. McGuinness. "Ontology development 101: A guide to creating your first ontology. 2001." See http://protege. stanford. edu/publications (2004).
6. Malik, Sanjay Kumar, Nupur Prakash, and S. A. M. Rizvi. "Ontology Creation towards an Intelligent Web: Some Key Issues Revisited." International Journal of Engineering and Technology 3, no. 1 (2011): 44.
7. van Rotterdam, Jeroen M., Michael Mohen, Ravi Ranjan Jha, and Sreecharan Shroff. "Memory centric database architecture." U.S. Patent 10,049,041, issued August 14, 2018.
8. Khobragade, P. V., and Nilesh Uke. "Cogent Sharing of Covert File Using Audio Cryptographic Scheme." International Journal of Applied Information Systems 1, no. 8 (2012): 1-4.
9. Beno, Martin, Erwin Filtz, Sabrina Kirrane, and Axel Polleres. "Doc2RDFa: Semantic Annotation for Web Documents." (2019).
10. Cressey, Gillian. Diaspora youth and ancestral homeland: British Pakistani/Kashmiri youth visiting kin in Pakistan and Kashmir. Brill, 2006.
11. Al-Azzeh J.S., Al Hadidi M., Odarchenko R., Gnatyuk S., Shevchuk Z., Hu Z. Analysis of self-similar traffic models in computer networks, International Review on Modelling and Simulations, № 10(5), pp. 328-336, 2017.
12. Odarchenko R., Abakumova A., Polihenko O., Gnatyuk S. Traffic offload improved method for 4G/5G mobile network operator, Proceedings of 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET-2018), pp. 1051-1054, 2018.
13. Z. Hassan, R. Odarchenko, S. Gnatyuk, A. Zaman, M. Shah, Detection of Distributed Denial of Service Attacks Using Snort Rules in Cloud Computing & Remote Control Systems, Proceedings of the 2018 IEEE 5th International Conference on Methods and Systems of Navigation and Motion Control, October 16-18, 2018. Kyiv, Ukraine, pp. 283-288.
14. M. Zaliskyi, R. Odarchenko, S. Gnatyuk, Yu. Petrova. A.Chaplits, Method of traffic monitoring for DDoS attacks detection in e-health systems and networks. CEUR Workshop Proceedings, Vol. 2255, pp. 193-204, 2018.