

Domain ontology proposition for central nervous system tumors in pediatric patients

Mariana C. Pereira^{1,2,3*}, Mell A. Matsuda^{1,2}, Lucas N. Lanferdini^{1,2}, Marialva Sinigaglia^{1,3} and Sílvia C. Cazella²

¹ Children's Cancer Institute, São Manoel Street, 850, 90620-110, Porto Alegre, RS, Brazil

² Federal University of Health Sciences of Porto Alegre, Sarmento Leite Street, 245, 90050-170, Porto Alegre, RS, Brazil

³ National Science and Technology Institute for Children's Cancer Biology and Pediatric Oncology – INCT BioOncoPed, Brazil

Abstract

Ontologies have been used to model the representation of knowledge, favoring information retrieval. In the health scenario, they can be an excellent alternative for solving gaps in information dispersion and lack of interoperability. This summary presents an ongoing project that aims to propose a domain ontology for the systemic visualization of oncopediatric patients with Central Nervous System tumors. Clinical records, clinical protocols and therapeutic guidelines are being screened to enrich the ontological model. The NeOn Methodology will guide the ontology construction processes, which will be modeled in Protégé. It is hoped that the model will contribute to the expansion of ontologies aimed at the pediatric public and the health sector.

Keywords

Central Nervous System, Health Information Interoperability, Ontology, Pediatric Oncology

1. Introduction

Childhood cancer represents only 3% of the population's total tumors [1], which creates the scenario that the disease is little explored in large-scale scientific studies. However, in the Brazilian pediatric scenario, cancer is the main cause of death in children and adolescents [1], which indicates the need for a careful look at the demands of this population. As a result, investments in pediatric oncology are increasing, which not only generates an increase in the chances of a cure but also in the amount of data available for study.

However, offering comprehensive and continuous care is a complex task that requires the collaboration of a multidisciplinary team, often geographically dispersed, as reference centers are mostly located in capital cities. This generates a large number of clinical records that are also scattered, which favors the lack of interoperability, standardization, and the difficulty of integrating data and clinical monitoring [2]. In addition, many terminological and classification systems were not developed for automation [3, 4].

In this context, ontologies emerge as a possible solution to the problems of interoperability, standardization and data recovery. In line with this, Ordinance No. 2.073/2011 of the Brazilian Ministry of Health encourages the use of common ontologies, terminologies and classifications

Proceedings of the 17th Seminar on Ontology Research in Brazil (ONTOBRAS 2024) and 8th Doctoral and Masters Consortium on Ontologies (WTDO 2024), Vitória, Brazil, October 07-10, 2024

* Corresponding author.

✉ marianacp@ufcspa.edu.br (M. C. Pereira); mell.matsuda@ufcspa.edu.br (M. A. Matsuda); lucas.lanferdini@ufcspa.edu.br (L. N. Lanferdini); msinigaglia@ici.org (M. Sinigaglia); silvioc@ufcspa.edu.br (S. C. Cazella)

ORCID 0000-0002-7938-8356 (M. C. Pereira); 0009-0005-0244-6852 (M. A. Matsuda); 0009-0005-3853-9794 (L. N. Lanferdini); 0000-0002-8324-6860 (M. Sinigaglia); 0000-0003-2343-893X (S. C. Cazella)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

for the definition and representation of concepts related to health [5]. Therefore, this project is based on the need to propose a domain ontology for the systemic monitoring and visualization of oncopediatric patients.

This article describes ongoing research to achieve the ontological proposition. The most relevant information is presented in five sections, including in Section 2 the theoretical framework, in Section 3 the related works, in Section 4 the research method and in Section 5 the final considerations. Due to the complexity of the topic and its incipient study, there was a special focus on cases of Central Nervous System (CNS) tumors.

2. Theoretical frameworks

2.1 Interoperability of Health Information and Ontologies

Interoperability of health information is understood as the ability of systems to exchange information and use it [6]. This exchange can be of a syntactic or semantic nature, and, in the semantic context, data agreement allows for more agile and automatic processing by computers [7]. Thus, interoperability and ontology complement each other, since the ontology acts as a guide to establish the standards that guide how terms relate to each other, acting as a terminological control that defines the language or set of terms used for queries in the domain area [8].

Regarding the development of ontologies, this is an interdisciplinary activity. Due to the ability to specify the semantics of terms in different domains and to establish semantic relationships between concepts, ontologies have been widely used in the health area [8-12], with a domain ontology being understood as being responsible for defining and characterizing the scope in which tasks will be performed, differentiating from task ontologies, whose focus is on solving specific problems [8].

2.2 The scenario of pediatric cancer and Central Nervous System tumors

Clinical oncology is the specialty dedicated to the study, planning, treatment prescription and monitoring of cancer patients [13]. Pediatric oncology serves patients aged 0 to 19 years and is related to the most diverse types of neoplasms, whether benign or malignant. In the scenario of this study, malignant tumors (popularly called cancer) will be treated and characterized by the disordered growth of cells [14].

Tumors that affect the child and adolescent population differ significantly from those that affect adults and tend to be more aggressive, resulting in more intense treatments and greater susceptibility to late effects [1]. In the case of Central Nervous System (CNS) tumors, they are the most common group of solid tumors in children and adolescents, accounting for 20% of all neoplasms in childhood and with a higher incidence in children aged one to four years [15].

Known as gliomas, astrocytomas, meningiomas, medulloblastomas, etc., the types of CNS tumors have different anatomical regions, classification, staging and treatment, as they present specific histological characteristics that will determine the pattern of dissemination of the disease [15]. Therefore, standardizing criteria such as staging and extent of disease is essential to allow comparison of performance between different national centers [16].

Opportunely, the increase in biomedical databases and the need for more sophisticated data retrieval favor the use of ontological models in healthcare and, therefore, improving terminologies, making them understandable to humans and processable by computers, has become an essential task. However, there are still gaps in the pediatric oncology domain, both in terms of health terminologies and the most appropriate models for building ontologies without practical ambiguities. For example, "tumor" can refer to both a physical mass and the illness process that the patient faces and, thus, there is the same word to characterize different entities,

as an entity is material - physical tumor with size and weight, and the other is a malignant disease – with duration in time [17].

Therefore, it is crucial to consider the complexity of domain details and the existence of ambiguities to propose ontologies that support semantic interoperability. Thus, when talking about standardization, we also talk about the development of ontologies. The relationship between pediatric oncology and ontologies will be presented below, in the explanation of related work.

3. Related work

A scoping review was performed to identify relevant work within the domain. The review and its details are registered in the Open Science Framework², with the works most aligned with the proposal of this project being presented. Although the area of application of ontologies in health is promising, the review revealed a scarcity of studies that specifically address the pediatric oncology scenario.

Santos et. al [18] propose an automated data mining system that uses ontology to facilitate access to analytical information about brain tumors by public health decision makers, especially in decentralized health systems. Similarly, Feilmester et al. [19] present a method for analyzing database attributes using a reference ontology to support the development of an information system for therapeutic planning of brain tumors, evaluating the effectiveness of the proposed method in comparison with traditional methods.

The paper of Stewart et al. [20] highlights the development of the YouCan Ontology framework, which uses a customized ontology to represent knowledge from a Clinical Practice Guideline and generate personalized self-management reports for pediatric patients in the post-treatment monitoring phase. The YouCan Ontology illustrates the ontology's ability to model medical knowledge and support shared decision-making between patients and healthcare professionals.

In another context, but still exploring the potential of ontology to integrate health data, Hansi et al. [21] propose an ontology-based semantic data integration framework for integrative analysis. The authors argue that the ontological approach facilitates the integration of heterogeneous data, in addition to improving the documentation and communication of integration processes.

The Protégé tool [22] stands out as a consensus among the studies reviewed for the development of ontologies. Various medical terminologies were utilized, including NCI Thesaurus, which offers standardized codes for biomedical concepts; SNOMED-CT, which provides organized medical terms and definitions for clinical documentation; ICD Codes, which classify diseases and their causes; TNM Classification, a global standard for malignant tumor staging; and UMLS, which integrates biomedical vocabularies to enhance system interoperability. Pre-existing ontologies, such as Disease Ontology, Ontology for Biomedical Investigations and Cell Line Ontology, were reused in some studies, indicating opportunities for reuse and collaboration in the area. However, the review also revealed a gap in the detailed description of the adopted ontology modeling methods.

4. Methods

The methods follow a structured set of methodological steps, the first of which is mapping variables from clinical records of pediatric cancer patients, which includes the identification and standardization of relevant clinical data, such as diagnoses, staging, and treatment of the disease.

² Domain ontology for pediatric oncology patients: a scoping review protocol - <https://osf.io/8a3ep/>

An ontological dictionary is in the developing stage to ensure consistency and standardization of information. The specification of terminological and conceptual relationships between domains will be a step that will ensure that the ontology accurately represents the complex interactions between different concepts. Finally, the validation of knowledge will occur through consultation with domain experts, consolidated sources of scientific literature, and the most widely used ontologies. The chosen methodology, NeOn, will be introduced in the next section).

4.1 Mapping data to ontology

Variable mapping was conducted using clinical records from an epidemiological study³ and made available by the Children's Cancer Institute (ICI). The clinical records, provided in blank format, only allowed the detailed identification and categorization of entries without involving patient information. Since no human data is involved, this study does not require consideration by the Research Ethics Committee, being registered and approved by the Research Committee of the Federal University of Health Sciences of Porto Alegre⁴ and by the ICI Research Projects Committee⁵.

Data structuring involved the creation of categories and subcategories that reflected the complexities and specificities of pediatric oncology cases. The main categories include demographic data, clinical characteristics, tumor specificities, treatment, responses to treatment and adverse effects, as well as psychosocial aspects and quality of life of patients. The categorization and organization of data are being oriented to support future clinical, epidemiological and public health analyses, aiming to provide a solid basis for the developed ontology.

4.2 Ontology modeling Ontology modeling

This study is in the development phase, employing the NeOn methodology [23] to model the initial structure of a CNS-focused ontology. NeOn provides nine flexible options to facilitate ontology and ontological network creation. For this work, scenarios 1, 2, and 9 were selected (Figure 1) which focus on the ontology lifecycle from conception to implementation. These scenarios encompass the reuse and reengineering of non-ontological resources and the localization of ontological assets, enabling the progression from draft to planned, conceptualized, and formalized ontology [23]. Given that this ontology is initially designed to address the terminological relationships in pediatric oncology, with a focus on health services and professionals, foundations such as BFO were not used in its construction. This is also one of the reasons why the aforementioned NeOn scenarios were chosen. Considering the main purpose of modeling, scenarios 1, 2 and 9 are the most suitable for our current needs.

³ "Acompanhar para transformar: um olhar integrado para o câncer infantojuvenil a longo prazo, considerando aspectos clínicos, psicológicos e sociais", CAAE 52044221.8.1001.5327

⁴ Number 201/2022

⁵ Number 2022-0108.01

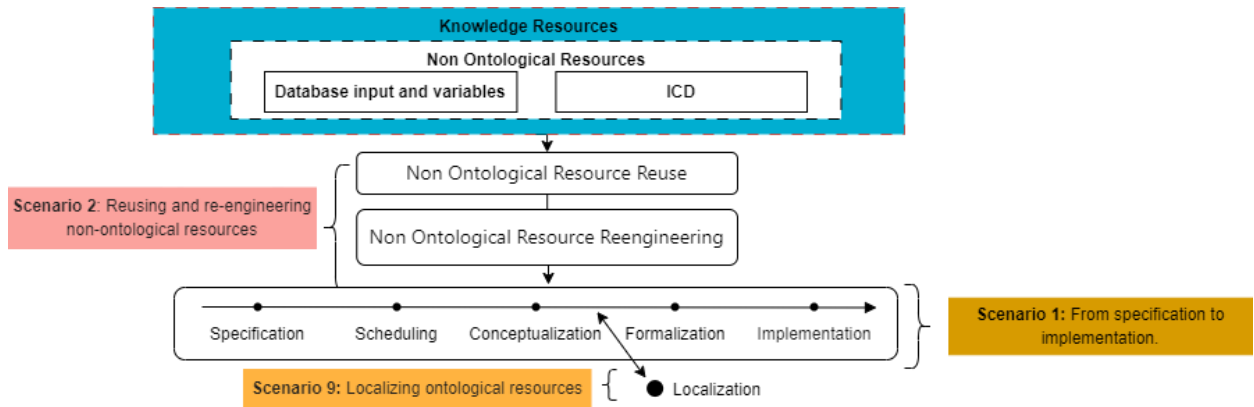


Figure 1: NeOn scenarios in use, prepared by the authors (2024).

Among the entities preliminarily mapped are the classification of tumors, diagnostic procedures, types of tumors, treatments performed, signs and symptoms and the international classification of the disease (Figure 2 and Figure 3). Entities related to the patient will also be included, such as gender, age group, place of birth, city and treatment center, etc. The codes presented in morphology and topography are the formal descriptions presented in the ICD-O [24] which allows the coding of all neoplasms by topography (site of disease), histology (morphology/type of disease) and biological behavior (disease staging).

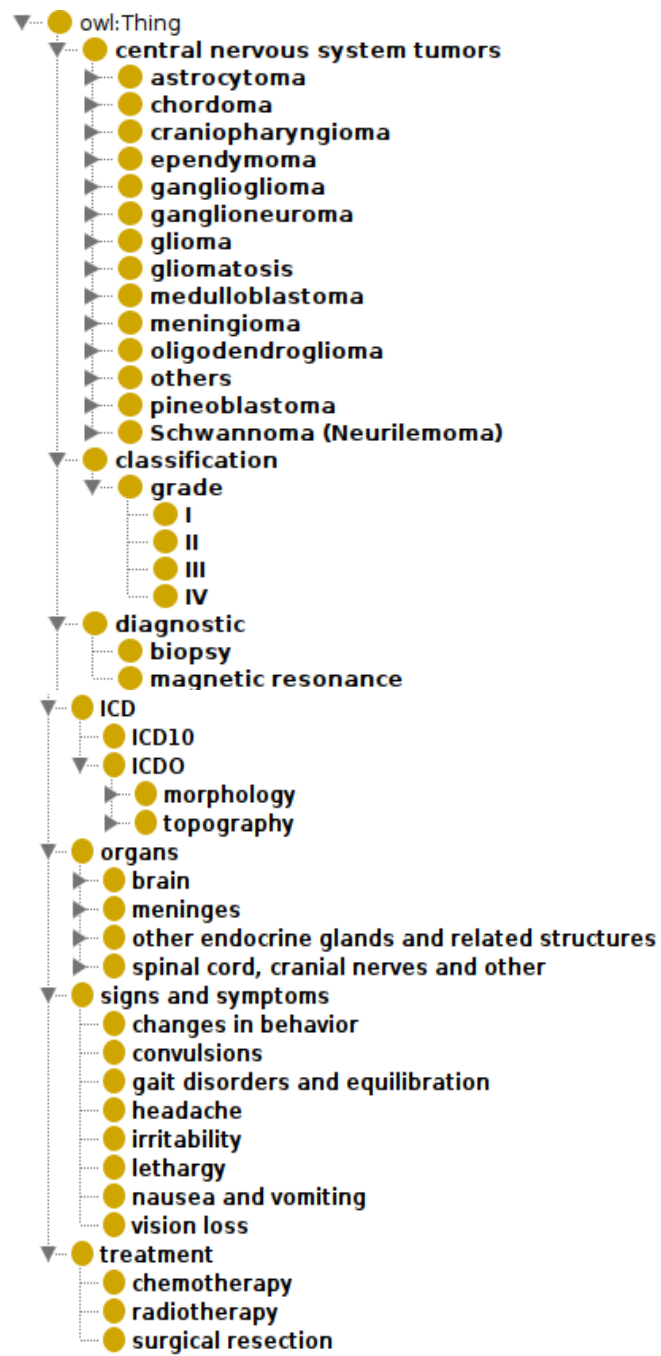


Figure 2: Entities modeling in Protégé, prepared by the authors (2024).

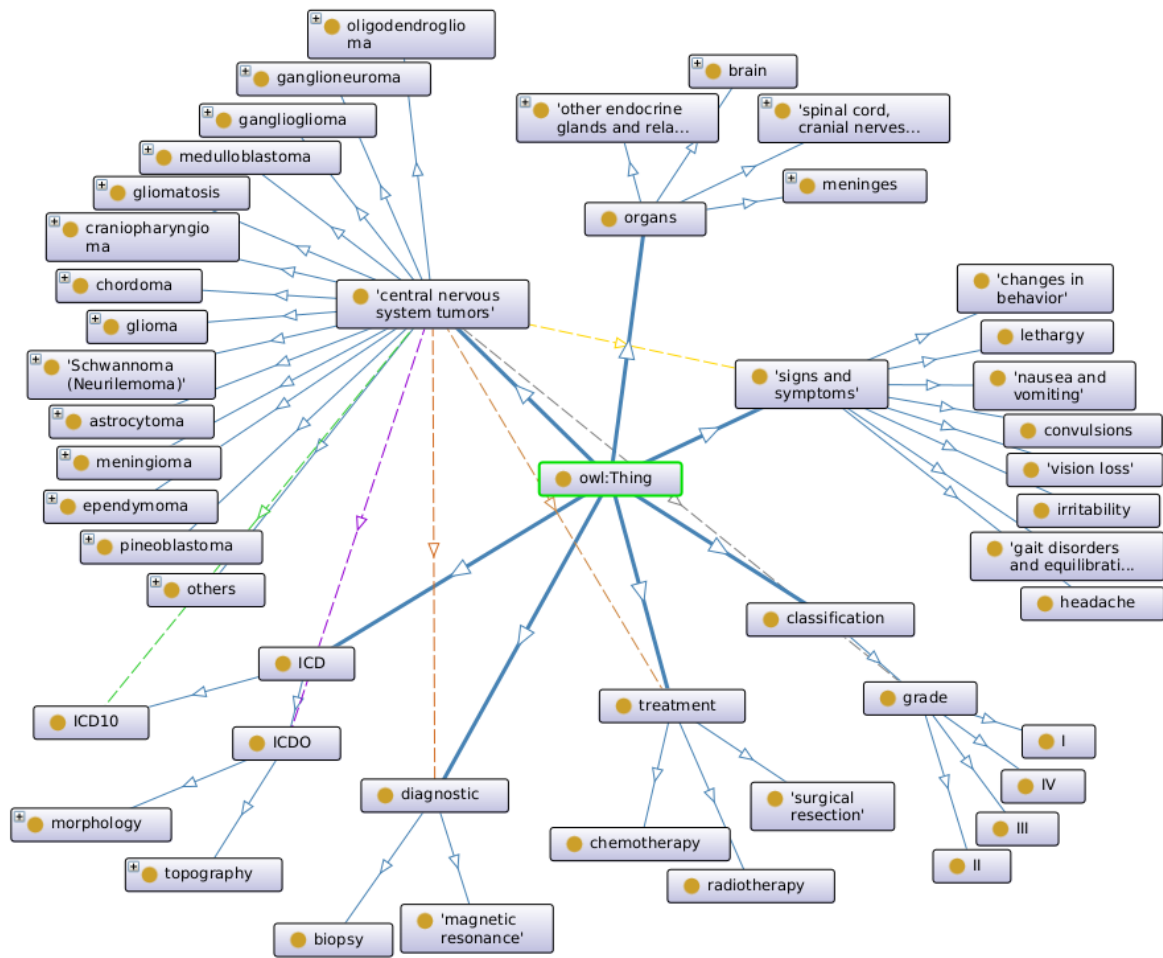


Figure 3: Preliminary modeling in Protégé, prepared by the authors (2024).

5. Final considerations

Regarding the study domain, ontology modeling could benefit clinical assistance, the development of scientific research and health education, however, ontology proposals for this domain are an incipient topic. Furthermore, the need to reduce the ontology domain for tumors of the CNS also presents itself as a limitation. The present work is in the final phase, and this article presented the proposal for a preliminary ontology modeling which aims to fill the gap highlighted. Considering that the ontological model has not been finalized, it is expected, in the next steps, that the ontology will be named, encompassing the scenario of pediatrics and central nervous system tumors. Also, it is expected that the proposed model can significantly contribute to the organization, mapping, understanding and interoperability of CNS tumor data for pediatric patients. The use of BFO is considered for future projects aimed at the computer science area.

Acknowledgments

State of Rio Grande do Sul Research Support Foundation, FAPERGS Project 22/2551-0000390-7 (RITE CIARS), Children's Cancer Institute and Ronald McDonald Institute, grant number 2022062.

References

- [1] National Cancer Institute - INCA, Childhood Cancer, Ministry of Health, 2018. URL: <https://www.inca.gov.br/tipos-de-cancer/cancer-infantojuvenil>
- [2] L. M. D. Teixeira, M. B., Almeida. Aspectos ontológicos e epistemológicos em terminologias clínicas: em busca de interoperabilidade semântica no ambiente clínico. *Encontros Bibli.* 2019 May 6;24(55):1–21. URL: <http://dx.doi.org/10.5007/1518-2924.2019.e57996>
- [3] A. Q. de Andrade. A linguagem médica utilizada em prontuários e suas representações em Sistemas de Informação: as ontologias e os modelos de informação. *Repositório Institucional da UFMG.* 2013 Nov 8; URL: <https://repositorio.ufmg.br/handle/1843/BUBD-9HHGLS>
- [4] A. Gómez-Pérez, M. Fernández-López, O. Corcho. *Ontological Engineering. with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web.* 1st Edition. London: Springer-Verlag, 2004.
- [5] Brazil. Ordinance No 2.073, August 31, 2011. Ministry of Health. URL: https://bvsms.saude.gov.br/bvs/saudelegis/gm/2011/prt2073_31_08_2011.html
- [6] IEEE. IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. *IEEE Std 610.* 1991;1–217.
- [7] Moreno RA. Interoperabilidade de Sistemas de Informação em Saúde. *Journal of Health Informatics.* 2016;8(3). URL: <https://docs.bvsalud.org/biblioref/2017/02/831874/editorial.pdf>
- [8] M. B. Almeida, M. P. Bax. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação.* 2003 Dec;32(3):7–20.
- [9] S. Isotani, I. I. Bittencourt. *Dados abertos conectados. Capítulo 3. Ontologias e Representação de Conhecimento.* São Paulo: Novatec. Editora, 2015. ISBN 978-85-7522-449-6.
- [10] Farinelli F, Melo S, Almeida MB. O papel das ontologias na interoperabilidade de sistemas de informação: reflexões na esfera governamental. 2013. URL: https://mba.eci.ufmg.br/downloads/Interoperab_Enancib_2013_camera-ready.pdf
- [11] Nilashi M, Ibrahim O, Bagherifard K. A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques. *Expert Systems with Applications.* 2018 Feb;92(2018):507–20. URL: <https://doi.org/10.1016/j.eswa.2017.09.058>
- [12] M. T. D. Melo, *Ontologia de arquivos multimídia como ferramenta de auxílio na classificação do câncer de mama.* 2019. 92 f. TCC (Graduação). Curso de Engenharia Eletrônica, Universidade de Brasília, Brasília, 2019.
- [13] R. Barroso-Sousa, G. Fernandes. *Oncologia: princípios e prática clínica.* 1. ed. Santana de Parnaíba [SP]: Manole, 2023.
- [14] National Cancer Institute - INCA. What is cancer? Ministry of Health. 2018. URL: <https://www.inca.gov.br/o-que-e-cancer>
- [15] National Cancer Institute - INCA. Central nervous system tumors. Ministry of Health. 2022. URL: <https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/infantojuvenil/especificos/tumores-do-sistema-nervoso-central>
- [16] Brazil. Law No. 14,308, May 8, 2022. *Official Gazette of the Union - Section 1, number 46, September 3, 2022.* URL: <https://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?data=09/03/2022&jornal=515&pagina=1>
- [17] Jansen, Ludger, and Stefan Schulz. "The ten commandments of ontological engineering." *Proceedings of the 3rd Workshop of Ontologies in Biomedicine and Life Sciences.* 2011.

- [18] Santos RS, Malheiros SMF, Cavaleiro S, Oliveira JMP de. A data mining system for providing analytical information on brain tumors to public health decision makers. 2013 Mar 1;109(3):269–82.
- [19] Felmeister AS, Waanders AJ, Mason JL, Stevens J, L. Charles Bailey, Ganesan S, et al. A Framework for Analysis, Ontological Evaluation, and Visualization in Preparation to Predictive Analytics in Pediatric Brain Tumor Research. 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2020 Dec 16;2020:1943-1950. URL: <https://10.0.4.85/bibm49941.2020.9313388>
- [20] Stewart SA, Abidi S, Parker L, Bernstein M, Abidi SSR. Clinical guideline-driven personalized self-management diary for pediatric cancer survivors. *Studies in Health Technology and Informatics*. 2014;2014(205):18–22. URL: <https://pubmed.ncbi.nlm.nih.gov/25160137/>
- [21] Zhang H, Guo Y, Li Q, George TJ, Shenkman E, Modave F, et al. An ontology-guided semantic data integration framework to support integrative data analysis of cancer survival. *BMC Medical Informatics and Decision Making*. 2018 Jul;18(S2). URL: <https://10.0.4.162/s12911-018-0636-4>
- [22] Stanford University. Protégé. 2019. URL: <https://protege.stanford.edu/>
- [23] Suárez-Figueroa MC, Gómez-Pérez A, Motta E, Gangemi A. *Ontology Engineering in a Networked World*. Springer eBooks. Springer Nature; 2012. URL: <https://link.springer.com/book/10.1007/978-3-642-24794-1>
- [24] World Health Organization. *International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3)*. 2019. URL: <https://www.who.int/standards/classifications/other-classifications/international-classification-of-diseases-for-oncology>