Domain Ontology Construction in the Glioma

Lanxin Huang, Na Sun*
School of Public Health, Shaanxi University of Chinese Medicine, Xianyang 712046, China.

Abstract: **Objective** To construct a domain ontology for glioma, and to achieve knowledge sharing and reasoning in the field of glioma through reuse and sharing of knowledge. **Methods** A domain ontology for glioma was constructed using the Ontology Web Language as the description language and a seven-step approach with the help of the ontology construction tool Protégé 5.5.0 as a knowledge source. **Results** A more comprehensive knowledge system related to glioma was presented, with 35 classes, 85 instances, 8 object properties and 5 data properties constructed. **Conclusion** By integrating the glioma knowledge and establishing the corresponding domain ontology, it provides knowledge support and reasoning basis for sharing, reusing and building the knowledge base of glioma knowledge.

Keywords: Glioma; Ontology; Protégé; Domain ontology

Introduction

Glioma is a tumour arising from cancerous changes in the glial cells of the brain and spinal cord. Among neurological tumours, it is one of the most common primary tumours [1]. It accounts for 50-60% of all primary cranial tumours, 20% of all neurological tumours and 80% of central nervous system tumours [2]. Studying the causes, diagnosis and treatment of glioma, it is important to establish a knowledge system for the integration of glioma knowledge and the construction of a knowledge base, which is scattered in medical books, journals, etc. It is important to construct the corresponding ontology of these multi-source heterogeneous data in a knowledge-based manner to prepare for the future knowledge sharing and application. Ontology has its origins in Western philosophy and is an inquiry into the origin of everything in the world [3]. The Semantic Web aims to achieve a very sophisticated management system by using automated tools to extract knowledge in the conceptual space [4]. At the heart of the Semantic Web's knowledge representation are ontologies that better describe the structure of knowledge and the specification of concepts at a semantic level, and that offer the possibility of knowledge sharing by discovering hidden relationships through logical reasoning.

Guarino [5] divided ontologies into three layers according to the degree of domain ontology dependency: top-level ontology, domain and task ontology, and application ontology, which are specialized ontology describing the relationship between concepts and concepts in a particular domain. Ontology have become a research hot spot, and with the continuous development of ontology technology, they are also gradually applied to knowledge expression in the medical domain, and are widely used in the biomedical field, such as Gene Ontology (GO) [6], and various ontologies for independent diseases, such as diabetes ontology, etc [7]. Most of these independent disease ontologies are constructed in terms of patient condition, disease staging, symptoms, and treatment, which can better assist in supporting clinical decision making. However, no brain glioma ontology has been found in single disease ontology studies. Our study is to provide ideas for other disease ontologies by building a glioma ontology that is designed to better integrate knowledge of glioma and allow knowledge from different sources to be brought together in a single body of knowledge to provide support for subsequent more comprehensive focus on the diagnosis, treatment and other processes of the disease.

1. Ontology construction

1.1 Ontology construction methods

There are various methods of ontology construction, and different methods produce different results. The common ones are the seven-step method, the skeleton method, the TOVE method, the METHONTOLOGY method, etc. The seven-step method was proposed by Stanford University School of Medicine in 1995 for domain ontology, and it is also the most appropriate ontology construction method for the medical field [8]. In this study, the fifth and sixth steps of the seven-step method were combined as follows: (i) identify the domain and scope of glioma; (ii) examine the possibility of reusing existing disease ontologies; (iii) list the important terms of the glioma ontology; (iv) define the hierarchy between glioma classes; (v) set the properties and their constraints; and (vi) create glioma instances.

1.2 Ontology building languages and tools

The Ontology Web Language (OWL) is one of the ontology description languages published by the W3C (World Wide Web Consortium) in 2004, developed from a combination of DAML and OLL, and builds on RDF and RDFs, and adds more semantics to the description logic. It has a more comprehensive and accurate description of semantics than RDFs and RDFs, is more powerful and functional, and is reasonably efficient for reasoning [9]. The Ontology Editor is used to build, develop and maintain ontologies. There are more than 60 ontologies building tools available, commonly used are Protégé, OntoEdit and others. Among these tools, the most widely used and popular ontology editing tool is Protégé [10] developed by the Stanford University Medical Intelligence Research Group. It is developed in Java and supports a wide range of ontology formats through various plug-ins, and has good openness and compatibility, making it the current tool of choice for ontology development. We used Protégé 5.5.0 to build the glioma domain ontology.

2. Glioma domain ontology construction

2.1 Defining the glioma domain and scope

The first step includes issues such as the scope and purpose of the domain. The purpose of the domain, i.e. the domain of glioma, is to structure glioma knowledge and build a more comprehensive knowledge map of glioma. Maja Hadzic et al. [11] proposed a four-dimensional generic ontology framework for the disease, including disease type, phenotype, treatment and etiology, which basically covers the most important features of the disease. we added two concepts to this top-level category, and the final brain The top level of glioma was identified as Basic_Information_of_Patient, Cause, Diagnosis, Sympton, Therapeutic_Procedure, and Tumor_Classification in a total of six areas. Knowledge is derived from encyclopedic knowledge, expert experience, and clinical guidelines including (National Comprehensive Cancer Network,NCCN) Clinical Practice Guidelines (2020.v3), the patient guide Brain Tumors - Glioma (v2021), (American Brain Tumour Association, ABTA), About Brain tumors-a Primer for patients and caregivers, and a literature search of glioma-related literature through literature databases.

2.2 Examining the possibility of reusing existing disease ontologies

Reusing ontologies can improve the efficiency of building ontologies and reduce the waste of resources. We searched BioPortal for relevant disease ontologies and reused terms from Thyriod Cancer Ontology, TCO and Brain Tumour Ontology, BTO.

2.3 List of important terms for glioma ontology

By manually extracting the entities related to glioma from the above knowledge sources, we obtained the concepts, properties and other contents of them, but in our process of collecting terms, there existed this term with multiple meanings, by referring to (NCI Thesaurus OBO Edition, NCIt), (systematized nomenclature of medicine-clinical terms, SNOMED CT) the relevant standard terminology was expressed in a standardized way whenever possible.

2.4 Defining a hierarchy between glioma classes

Based on the conceptual terms generated from the above steps, the analysis was followed by hierarchical relationship building. The core of the domain ontology is the class, and the hierarchical relationship between classes is the hierarchical relationship between concepts, using a combination of top-down and bottom-up approaches. The top-level concepts have six major classes which are Basic_Information_of_Patient, Cause, Diagnosis, Sympton, Therapeutic_ Procedure, and Tumor_Classification.

2.5 Defining properties and facets of classes

The properties of ontology classes include Object property and Data property. Object property represents the relationship between classes and classes, while data property represents the relationship between instances and data value types. Protégé has its own basic relationships including part-of, kind-of, instance-of, attribute-of. We have created a total of eight Object properties, defined by range and domain, which allow classes to be related to each other by means of this property, laying the foundation for the application of knowledge inference. For example, the Object property "treatedBy" relates Anaplastic_oligodendroglioma in "Grade III" to Chemotherapy in "Chemotherapy". Chemotherapy_Administered_During_Radiation_Therapy" in "Chemotherapy". A total of 5 Data attributes are created including characterize, has_ICDO-3_Code, has_SNOMED_CT®_ConceptId, hasMeshID and requireCOST, e.g. "has_ICDO-3_Code " and "has_SNOMED_CT®_ConceptId" have Functional properties because different gliomas can only have one International Classification of Diseases for Oncology or SNOMED_CT®_ConceptId. Diseases for Oncology or SNOMED_CT® with the property value xsd:string. e.g. Pilocytic_astrocytoma has_ICDO-3_Code "9421" and has_SNOMED_CT ®_ConceptId "128854008".

2.6 Creating glioma instances

After building the upper ontology and defining the properties, the final step requires the creation of Individuals, instances are concretizations of concepts i.e. individuals, which cannot be subdivided and inherit the properties of the class. Individuals are set in the symptoms, such as Seizures, Fatigue, Double_vision, etc., considering the existence of concretization properties for terms related to symptoms (Sympton) of glioma.

3. Reasoning and testing of glioma ontology

The consistency of domain ontologies includes three types of consistency: syntactic consistency, semantic consistency and consistency of user-defined domain rules [12]. After the ontology is constructed, it is first manually verified and then reasoned by Protégé 5.5.0's own reasoning machine HermiT 1.4.3.456, an OWL-based ontology reasoning machine developed by the Computer Science Data and Knowledge Research Group at the University of Oxford, which finds hidden potential by the consistency of our previously defined relational properties and axioms relations and instances. The results of the inference are presented in pale yellow and if the wrong relationship is established, the system will report an error for you to make changes. As one of the glioma symptoms, Headache can then be automatically categorised by the reasoning machine, as shown in Figure 1. Based on the semantic relations we have constructed, diseases such as Glioblastoma and Optic_nerve_glioma have Headache as a symptom, and the reasoning results will show what diseases have Headache as a symptom. The ontology we have constructed has passed the consistency test and has internal logical consistency.

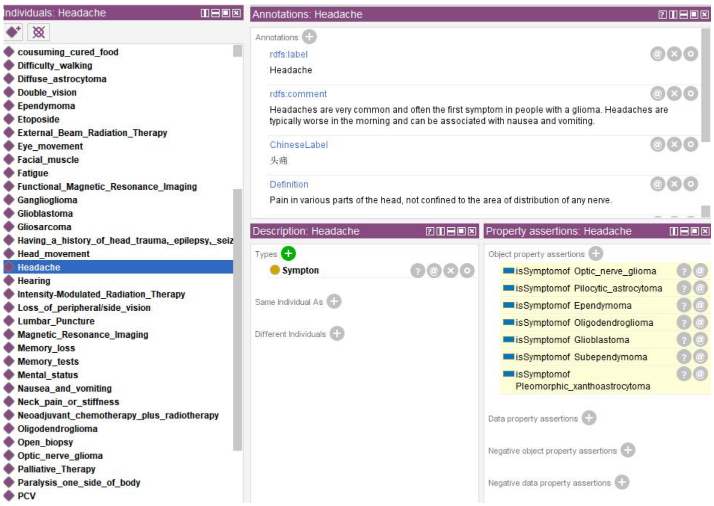

Figure 1. Example of ontology consistency reasoning

4. Graphical presentation of the glioma ontology

The glioma ontology constructed in this paper has 35 classes, 85 instances, 8 object properties and 5 data properties. This knowledge ontology is graphically presented using OntoGraf, a tool that comes with Protégé (Figure2). This figure shows the structure of the glioma ontology, with nodes representing semantic concepts, solid lines representing relationships between classes, and dashed lines representing different object properties connected by arrows, with different arrow colours representing different semantic relationships.

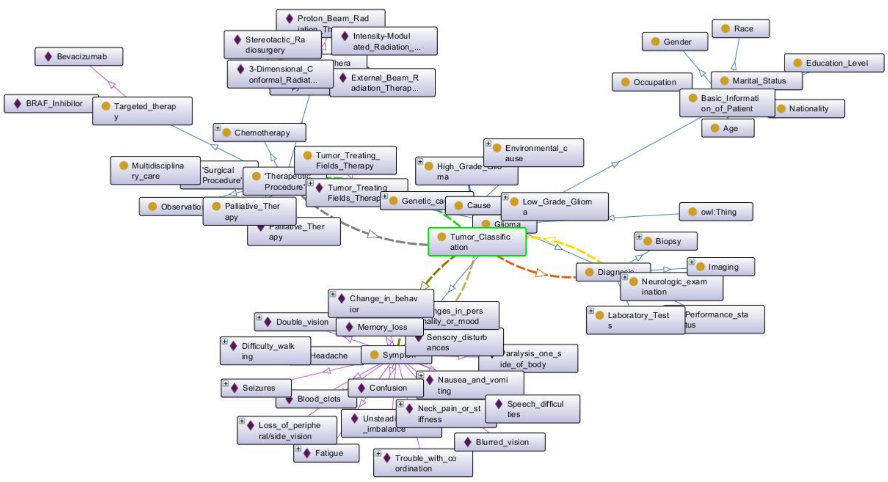

Figure 2. Structure of the glioma proper

Conclusion

A domain ontology is a description of the concepts and specific relationships between concepts in a relevant professional domain, which allows for a more effective and rational representation of knowledge. The combination of medicine and ontology allows for more focused knowledge exploration and relationship discovery around glioma, which not only facilitates knowledge sharing and application, but also provides a reference for other disease ontologies. However, this ontology has some limitations, mainly in the following areas. Because the ontology was constructed manually, the constructed glioma domain ontology may not be complete and more domain experts are needed to participate in it. Future work will require more in-depth research and extensions, such as automated methods for building ontologies, adding conceptual terms related to glioma-related aspects, or building SWRL rule bases for reasoning.

References

[1] Ostrom QT, Gittleman H, Stetson L, Virk SM, Barnholtz-Sloan JS. Epidemiology of gliomas. Cancer Treat Res. 2015;163:1-14.

[2] Dolecek TA, Propp JM, Stroup NE, Kruchko C. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2005-2009. Neuro Oncol. 2012 Nov;14 Suppl 5(Suppl 5):v1-49.

[3] Li Shanping, Yin Qiwei, Hu Yujie, Guo Ming, Fu Xiangjun. A review of ontology research[J]. Computer Research and Development,2004(07):1041-1052.

[4] Antoniou G, Van F H . A Semantic Web primer[M]. MIT Press, 2004.

[5] Guo HY. Study on the construction of ontology model of disease domain [D]. Chinese People's Liberation Army Academy of Military Medical Sciences, 2011.

[6] Ashburner M, Ball C.A, Blake J. A, et al. Gene ontology: tool for the unification of biology[J]. The Gene Ontology Consortium. Nature Genetics, 2000, 25(1): 25-29.

[7] Shaker El-Sappagh, Farman Ali. DDO: a diabetes mellitus diagnosis ontology[J]. Applied Informatics, 2016, 3(1).

[8] Chen ZJ. Research on ontology-based knowledge retrieval system for diagnosis of digestive system diseases [D]. Chongqing University,2017.

[9] Yuan Y. Research on the construction of pathological ontology for colorectal cancer[D]. Zhejiang University, 2016.

[10] Rubin DL，Noy NF，Musen MA. Protege: a tool for managing and using terminology in radiology applications. Journal of Digital Imaging, 2007, 20 (Suppl 1): 34-46.

[11] Hadzic M, Chang E. Ontology-Based Support for Human Disease Study[C]// Hawaii International Conference on System Sciences. IEEE, 2005.

[12] Xu Y, Wang ZX, Li ZY. Consistency checking of domain ontologies[J]. Computer Engineering, 2009,35(01): 55-57+64.