1.0 Introduction

Domain knowledge is important in business understanding and data understanding for implementing artificial intelligence solutions for any business problem. Ontology is a formal explicit description of concepts in a domain of discourse (classes or concepts), properties of each concept describing various features and attributes of the concept (slots or roles or properties), and restrictions on slots (facets or role restrictions). Ontology together with a set of individual instances of classes constitutes a knowledge base. It provides a common vocabulary to share information among people or software agents (Noy & McGuinness, 2001).

Domain specific ontology based search engines are more efficient than keyword based searches as semantic and contextual information are not captured in the extraction of keywords from the text (Malik et al., 2015).

2.0 Ontology development

Malik et al. (2015) had proposed a methodology using the tool Protégé to develop the ontology for agriculture domain. The proposal was based on the 7 steps described by Noy & McGuinness (2001) which are:

Step 1: Determine the domain and scope of the ontology
Domain: Agriculture, no generic ontology available at that time of writing
Scope: Use as a tool for text mining applications, to represent semantic relationships between entities to overcome limitations of keyword based search
Step 2: Consider reusing existing ontologies
Resources available: AGROVOC, multilingual thesaurus of food and agriculture
Step 3: Enumerate important terms in the ontology
Important terms in agriculture domain were extracted
Step 4: Define the classes and the class hierarchy
Five main classes identified: Plant, Disease, Pest, Pesticide and Fertilizer
Step 5: Define the properties of classes – slots
Properties (subclasses) added with respect to each class. Ex: For class Plant, the subclasses Cereal, Fruit, Vegetable and Pulses defined
Step 6: Define the facets of the slots
Define what value the slot can and cannot have
Step 7: Create instances
Create most specific concepts for each sub-class. Ex: Finger millet, Foxtail millet, Proso millet and Pearl millet are the instances for the subclass Millet

The above model was proposed to be created as a generic approach in preliminary stages. It would have missed to define some other classes or sub-classes, which could be added at a later stage, but the important thing is to get the hierarchy of the classes correct at any stage.

3.0 Ontology Application areas

3.1 Radiation Oncology

Radiation Oncology involves radiation generating equipment, multi-disciplinary team and quality management system to deliver radiation dose to the tumour for cancer patients. There are so many parameters and metrics that the treatment outcome depends on and the quantification and data sharing between different institutions and countries are very important to develop better treatment outcome models. Also developing new technologies such as individualised medicine, machine learning, and genetic profiling requires collecting big data of high quality. Ontologies can be a key element in the effort to achieve these goals.

3.2 Digital Library

In research. literature review is a key process to finding solution to the research question and takes much of the research time. Because of the deficiency of comprehensive disclosure of the characteristics of resources, often due to the inconsistency between users and the concept expression in the resource library, many resources remain out of reach for the researchers (Xuanhui & Fan, 2016). Hence systematic organisation of the library resources is required to connect the resources for better visualisation to researchers. Digital library ontology should be embraced to describe resources, and then the relationship between domain ontologies should be acquired by linguistic and statistical methods to make resources interconnected and shared (Shiyan, 2015)

4.0 Rationale for the approach used

There are many ways to create ontology model for a domain. Noy & McGuinness (2001) had proposed a simple knowledge-engineering methodology for developing ontology, which is an iterative approach. The modelling could start with a rough first pass and then along the way based on the requirement, pros, cons and implications of different solutions, the model could be refined.

In radiation oncology, efforts are being made to standardise vocabularies, establish taxonomies and interoperability standards (Philips, 2020). Developing ontology through the proposed approach could accomplish these goals and avoid differing standards.

In digital library, classification of resources may be based on language, subject, resource type, author, etc which would be a suitable domain for ontology deployment. Li, (2020) had presented a proposal for electronic book ontology model, with the ontology constructed from the perspective of class, relation and instantiation of ontology.

5.0 Feasibility of the proposed approach

Based on the ontology model approach proposed by Noy & McGuinness (2001)**,** ontologies models for the domain radiation oncology and digital library could be created as follows

5.1 Radiation Oncology

Step 1: Domain: Radiation Oncology; scope: To extract clinical data in standardised format in terms of nomenclature, structure and content for machine learning applications to be used by researchers to evaluate the treatment outcome

Step 2: Consider reusing existing ontologies: Could reuse the previously developed ontologies for radiation oncology such as:

- the Dependency Layered Ontology for Radiation Oncology (https://bioportal.bioontology.org/ontologies/DLORO)
- the Radiation Oncology Ontology (https://bioportal.bioontology.org/ontologies/ROO)
- the Radiation Oncology Structures Ontology https://bioportal.bioontology.org/ontologies/ROS).

Step 3: Enumerate important terms in the ontology: Important terms in Radiation Oncology could be listed, such as: Diagnosis, histology, Prescription, treatment fractionation, organ dose, tumour dose, etc. This could be expanded at later stage.

Step 4: Define the classes and class hierarchy: The following are example of class and sub-classes [Class: prescription; Sub classes: Fractionation, Technique]

Step 5: Define the properties of classes-slots: The sub-class: Fractionation has the slots [Total dose, Number of fractions]

Step 6: Define the facets of the slots: Slots have different facets such as the value type, number of values and any value the slot can take. [Ex: Total dose 'number', Number of fractions 'number']

Step 7: Create instances: Creating instances of classes in the hierarchy [Total dose: 60 Gy, Number of fractions: 30]

5.2 Digital Library

Step 1: Domain: Digital Library; scope: To construct book ontology such that the literature attributes could be exported to other data formats for convenient use in research systems for referencing.

Step 2: Consider reusing existing ontologies: Could reuse the previously developed ontologies for literature, if any.

Step 3: Enumerate important terms in the ontology: Important terms in literature domain would be listed, such as: book, author, chapter, publisher, etc

Step 4: Define the classes and class hierarchy: The following are examples of class and sub-classes [Class: publisher; Sub classes: name, date published]

Step 5: Define the properties of classes-slots: The sub-class: date published has the slots [Month, Year]

Step 6: Define the facets of the slots: Slots have different facets such as the value type, number of values and any value the slot can take. [Ex: Month: 'string' , Year: 'number']

Step 7: Create instances: Creating instances of classes in the hierarchy [Month: May, Year: 2022]

Once these preliminary models are created, it could be further developed based on the intended application and upon discussion with domain experts and end users

6.0 Analysis of the proposed approach

While developing ontology model, defining classes and hierarchy, there is no single correct model (Noy & McGuinness, 2001). The model could be made as per its application and level of required use. Once an initial model is created, it has to be evaluated by testing its application and discussing with the domain experts to iteratively revise and improve the model.

Modelling errors occur if incorrect hierarchy is followed (Noy & McGuinness, 2001):

- Correct hierarchy should be followed between classes – subclasses
- Either singular or plural form should be used in naming classes
- As domains evolve with successive iteration, sustaining a stable class hierarchy may become challenging
- Modelling error is possible if a class has only one direct sub-class.
- In cases where there are more than 10 subclasses for a class, it may be required to have some intermediate classes

Both extremely nested hierarchy classes as well as very plane hierarchy with few classes with more data in the slots are not optimal

The classes, subclasses, slot names has to be strictly followed to be consistent to improve the readability of the names.

In Radiation Oncology, the data is very large and unifying the data convention across the entire community is a big and expensive project. Also the equipment provided by different vendors follows different nomenclature and collating data between these systems in different data formats is a difficult problem. One solution is that the community should get together with scientific work groups to develop unified nomenclature and work towards its implementation across all systems.

Aldosari et al., (2017) had reviewed the negative sides of ontology in the field of medicine, that ontologies are perceived as a mere tool to represent medical knowledge and may not suffice the real-world scenario of dealing with complex patient records, which are grammatically complex, and documented in many native languages.

There is equally big problem in forming unified class hierarchy in digital libraries too. The digitisation of print books processed by scanning, limits the reconstruction of full-text and hinders key-words extraction. Also extraction of resources in different languages would be a difficult task too

7.0 Conclusion

The ontology development approaches was discussed and its application areas were put forward. Following the modelling guidance, any two ontologies created for same domain may not be the same. Ontologies should evolve based on its application and input from expert in domain knowledge.

8.0 References

Malik, N., Sharan, A. & Hijam, D. (2015) Ontology development for agriculture domain. *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)* 738-742.

Noy, N. F. & McGuinness, D. L. (2001) Ontology Development 101: A Guide to Creating Your First Ontology. *Knowledge Systems Laboratory.*

Aldosari, B., Alanazi, A., & Househ, M. (2017) Pitfalls of Ontology in Medicine. *Studies in health technology and informatics 238*: 15–18.

Philips, M. H. et al. (2020) Ontologies in radiation oncology. *Physica Medica* 72: 103-113.

Li, X. (2020) Study on book ontology in the digitisation process of books *Journal of Physics: Conference Series* 1544 DOI: https://doi.org10.1088/1742-6596/1544/1/012035

Shiyan, O. (2015) Semantic Representation of Chinese Thesauri. *Library and Information Service* 16: 110-118.

Xuanhui, Y. & Fan, L. (2016) Study on library digital resource service model based on ontology learning. *Research on Library Science* 9: 90-94