Ontology Patterns for Tubular or Spherical Layered Structures. A Case Study from Oncology

Johannes HERRMANN [a,b1], Susanne ZABKA [b], Martin BOEKER [b], Stefan SCHULZ [c]
[a] *Department of Computer Science, University of Freiburg, Germany*
[b] *Institute for Medical Biometry and Medical Informatics & Comprehensive Cancer Center, Faculty of Medicine and Medical Center, University of Freiburg, Germany*
[c] *Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria*

Abstract. The TNM classification supports malignant tumour staging and classification, based on size and location. The ontology framework TNM-O was recently proposed as a formal-ontological foundation of TNM. A number of gastrointestinal tumours have similar rules for classification, dependent on the anatomical layers of the wall of the digestive tract the tumour invades into. In this paper we propose a generalized pattern to represent spatial relations between the layers of tubular or spherical structures in anatomy. Using this pattern, we can create a strict total order on the layers, which proves useful for reasoning about the location of entities like tumours that spread across gastro-intestinal wall structures.

Keywords. ontology, layered structures, mereotopology, TNM classification, gastrointestinal tumours

1. Introduction

The TNM system is the most important coding scheme for malignant tumours. It is based on three components:

- T - The extent of the primary tumour
- N - The absence or presence and extent of regional lymph node metastases
- M - The absence or presence of distant metastases

The addition of numeric values indicates the extent of the malignant disease, e.g. T0, T1, T2, T3, T4 [1], for the extent of the primary tumour in a tissue, on which we focus in this paper. T1 to T4 are proportional to the size of the tumour, dependent on the organ in which it evolves, and the anatomical layers it affects. E.g. a tumour found in the innermost layer only is a T1 tumour, whereas a tumour that invades into the outermost layer and neighbouring body parts is classified as T4. T2 and T3 are assigned if the tumour was found in one of the middle layers.
Although all gastrointestinal tumours can be coded according to these rules, systematic

[1]Corresponding Author: Johannes Herrmann, Institute for Medical Biometry and Statistics, Medical Center – University of Freiburg, Faculty of Medicine, Stefan-Meier-Str. 26, 79104 Freiburg i. Br., Germany; E-mail: johannes.r.herrmann@gmail.com.

differences in the coding of the affected intestinal segment and layer are problematic. TNM-O is an ontological framework for the TNM classification system, which can be used to assign TNM classes to real world pathology data [2]. With the current approach, there is an ontology part defined for each tumour localisation, which represents the extent of the tumour. In a previous study [2], T1 to T4 were distinguished based on the infiltration depth of the layers. In the following class or individual descriptions of tumours, the feature "**btl2:isBearerOf** some (*Confinement* and (**btl2:projectsOnto** some *Invasive*))" is mentioned. This feature is used to differentiate between TNM codes which are not discussed in this paper. It was only included as part of a citation of [2] and has no effect on the application of the pattern.

> *InvasiveTumorOfSubmucosaOfColonAndRectum*
> EquivalentTo *ColonAndRectumTumour* and
> **btl2:isBearerOf** some (*Confinement* and (**btl2:projectsOnto** some *Invasive*)) and
> **btl2:isIncludedIn** some *SubmucosaOfLargeIntestine*

In cases where the tumour crosses multiple layers of the wall of the digestive tract, however, this definition is not unambiguous. Consider an individual tumour with the following features:

- *ColonAndRectumTumor*
- **btl2:isBearerOf** some (*Confinement* and (**btl2:projectsOnto** some *Invasive*))
- **btl2:isIncludedIn** some *SubmucosaOfLargeIntestine*
- **btl2:isIncludedIn** some *MuscularisPropriaOfLargeIntestine*

The HermiT reasoner [5] classifies this individual as both a T1 and T2 tumour, even though those classes should be disjoint. This problem can be solved by specifying exactly in which layers the tumour may or may not be included:

> *InvasiveTumorOfSubmucosaOfColonAndRectum*
> EquivalentTo *ColonAndRectumTumor* and
> **btl2:isBearerOf** some (*Confinement* and (**btl2:projectsOnto** some *Invasive*)) and
> **btl2:isIncludedIn** some *SubmucosaOfLargeIntestine* and
> **btl2:isIncludedIn** some *MuscularisPropriaOfLargeIntestine* and
> not (**btl2:isIncludedIn** some *SubserosaOfLargeIntestine*) and
> not (**btl2:isIncludedIn** some *VisceralPeritoneumOfLargeIntestine*)and
> not (**btl2:isIncludedIn** some (*Organ* and not (*LargeIntestine*)))

TNM distinguished the following anatomical locations for gastrointestinal wall tumours, viz. oesophagus, stomach, small intestine, appendix and colon/rectum. At least 25 complex class definitions of this form are needed for a full representation of TNM codes for all of the above anatomical locations and at least five layers. This would add some unnecessary redundancy and complexity to TNM-O. The objective of this work is therefore to provide a generalising pattern, for more concise definitions that classify all tumour configurations correctly. The pattern presented in section three is the main result of this work.

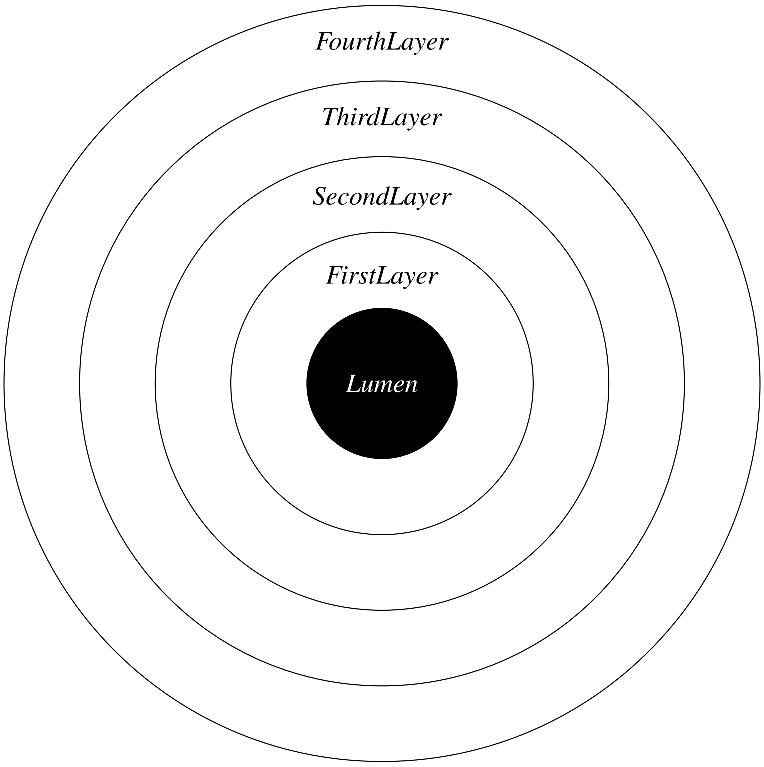

Figure 1. Cross section of a tube- or sphere-like object with $n = 4$ layers. In our example, the FirstLayer corresponds to the submucosa, the SecondLayer corresponds to the muscularis propria, the ThirdLayer corresponds to the subserosa, the FourthLayer corresponds to the visceral peritoneum and the FifthLayer corresponds to adjacent structures. See section 3.2

2. Methods

The ontology was created using Protégé [3] under the domain top-level ontology BioTopLite2 [4]. The HermiT DL reasoner [5] was used for classifying the individuals.

3. Results

3.1. The Pattern

The aim of the pattern is to generalize relationships between layers of a wall structure. This structure has to be spherical or tubular, such that each layer either is completely surrounded by or completely surrounds some other layer. In addition, the number of layers has to be finite. Let n denote that number of layers. For simplicity, we will refer to the innermost layer as the first one, and to the outermost one as the n-th layer (cf. Fig. 1).
In order to model this tubular, layered structure, a superclass for all of the structure's components is created. In our example [6], it is named *LayeredStructureComponent*. Its subclasses are classes of layers, of which complete tubular structures are composed. In order to formulate OWL axioms for the classes that represent the layers, two transitive

object properties are required, viz. **isSurroundedBy** and its inverse **surrounds**. This reflects our understanding that the fourth layer surrounds not only the third layer, but also the second and the first one.

The property **isSurroundedBy** has a subproperty chain: **btl2:isPartOf** ∘ **isSurroundedBy**. Thus, an object which is part of some layer is surrounded by the same entities as the layer itself.
The general pattern for the *i*-th layer, with $1 \leq i \leq n-1$, is:

> *i-thLayer* EquivalentTo
> *LayeredStructureComponent* and
> not (**isSurroundedBy** some *i-thLayer*) and
> **isSurroundedBy** some $i+1$-*thLayer*

The *n*-th layer has a slightly different definition:

> *n-thLayer* EquivalentTo
> *LayeredStructureComponent* and
> not (**isSurroundedBy** some *n-thLayer*)

By definition of these classes, the property **isSurroundedBy** forms a strict total order on the layers. The relation is transitive by definition and trichotomous, because for each two layers they are either the same or one surrounds the other. We hypothesise that this feature is useful once the pattern is applied to TNM-O.
To make this pattern more comprehensible for humans, we introduce two more classes, which do not change the semantics of the pattern: the *InnerLayer*, which in our example is equivalent to the *FirstLayer* and the *OuterLayer*, which is equivalent to the *n-thLayer*. Thus, independent of the application, it is obvious in which order the layers were enumerated. Furthermore, it is possible to use the definition of the *n-thLayer* as a general definition for the *OuterLayer*, because of its independence from the number of layers.

3.2. Application in TNM-O

We will use colorectal tumours as an example for the application of this pattern. The application to other tumour types mentioned in the introduction is similar. The only differences are the names of the layers and tumours.
In the seventh edition of the TNM the primary tumour is classified as follows [1]:

- T1 Tumour invades submucosa
- T2 Tumour invades muscularis propria
- T3 Tumour invades subserosa or into non-peritonealized pericolic or perirectal tissue
- T4 Tumour directly invades other organs or structures and/or visceral peritoneum
  - T4a Tumour perforates visceral peritoneum
  - T4b Tumour directly invades other organs or structures

The crucial information for classifying the primary tumour is its depth of invasion: the deeper the tumour invades the wall from the lumen of the gut, the higher it is classified in TNM. In other words: the submucosa of the colon is surrounded by the muscularis

propria, the muscularis propria is surrounded by the subserosa and so on. The last "layer" can be envisioned as the surrounding organs. This is not a layer in the conventional way, but these organs and structures are only relevant as they surround the visceral peritoneum. So we can view the mereological sum of the surrounding structures as layer in order to apply the pattern.
To do so, we simply rename the layers according to the classification:

FirstLayer → *Submucosa*
SecondLayer → *MuscularisPropria*
ThirdLayer → *Subserosa*
FourthLayer → *VisceralPeritoneum*
FifthLayer → *AdjacentStructure*

For the definition of gastrointestinal tract tumours it is important to consider that they generally grow outwards. For example a T3 tumour may not only invade into the subserosa, but also the muscular layer and the submucosa. To model this we will use the strict total order described above. The class definitions for the T1 and T2 tumours are as follows:

InvasiveTumorOfSubmucosaOfColonAndRectum
EquivalentTo *ColonAndRectumTumor* and
btl2:isBearerOf some (*Confinement* and (**btl2:projectsOnto** some *Invasive*)) and
not (**btl2:isPartOf** some *MuscularisPropria*) and
btl2:isPartOf some (**isSurroundedBy** some *MuscularisPropria*)

InvasiveTumorOfMuscularisPropriaOfColonAndRectum
EquivalentTo *ColonAndRectumTumor* and
btl2:isBearerOf some (*Confinement* and (**btl2:projectsOnto** some *Invasive*)) and
not (**btl2:isPartOf** some *Subserosa*) and
btl2:isPartOf some (**isSurroundedBy** some *Subserosa*)

The definitions for the other tumours are similar: aside from confinement and invasiveness, which are the same for the T1 to T4b tumours, they differ only in the layer in which they are not located. By using the **isSurroundedBy** property, the model represents that the tumour may be a part of all layers surrounded by the layer that must not contain the tumour.
These definitions are much more concise than those shown in the introduction, while still correctly classifying the tumours. Consider for example the following individual, which represents a T3 tumour:

- *ColonAndRectumTumor*
- **btl2:isBearerOf** some (*Confinement* and (**btl2:projectsOnto** some *Invasive*))
- not (**btl2:isPartOf** some *Subserosa*)
- **btl2:isPartOf** some *Submucosa*
- **btl2:isPartOf** some *MuscularisPropria*

This tumour is correctly classified by the reasoner [5] (see T3_Tumor individual in the example file [6]), even though it is part of two layers, like the individual shown in the introduction.

Another feature of this pattern is that the layer in which the tumour is not part of is important for the classification. Assume the tumour described above would not have the characteristic ”not (**btl2:isPartOf** some *Subserosa*)”. In this case the individual could be in the subserosa or not, which means that the tumour could also be a T3, T4a or T4b instead of a T2. Because of this uncertainty, the reasoner does not classify that individual, which is correct.

4. Conclusion and Outlook

Many structures of organisms share commonalities in their basic construction. One important example of such a structure is the hollow organ, which can be composed of several layers. Pathological processes may extend across these layers.

We proposed a generalization of the description principles of these processes and structures by engineering a general ontology design pattern. We demonstrated [6] that based on simple axiomatic assumptions an OWL-classifier drew correct conclusions.

However there are certain limitations to this pattern. It does not provide a distinction between the first layer and the lumen. This problem arises, because parthood and containment are difficult to delineate, especially in the biomedical context as discussed in [7]. Consider for example food in the digestive tract. One could argue, that this food is not part of the digestive tract, because before ingestion it was a separate entity. On the other hand, during digestion a number of nutrients are absorbed in the digestive tract and become part of the body. In order to clearly discern between lumen and first layer, a pattern or model is needed which deals with these problems.

As much as the proposed pattern simplifies the modelling of anatomical entities within the TNM ontology, the correct place for it would be an ontology of anatomy proper, like the Foundational Model of Anatomy (FMA). The FMA had introduced useful abstractions for anatomical entities [8], but it has not used the full descriptive power of OWL. The integration of the proposed pattern into a redesigned version of the FMA (or another anatomy ontology) would be desirable.

References

[1] Sobin, L. H., M. K. Gospodarowicz, and C. Wittekind. ”TNM classification of malignant tumours 7th edn Wiley.” (2009).

[2] Boeker, M., França, F., Bronsert, P., & Schulz, S. (2016). TNM-O: ontology support for staging of malignant tumours. Journal of biomedical semantics, 7(1), 64.

[3] Musen, M.A. *The Protégé project: a look back and a look forward.* AI matters 1.4 (2015): 4.

[4] Schulz, S., Boeker, M., & Martinez-Costa, C. (2017). The BioTop family of upper level ontological resources for biomedicine. Stud Health Technol Inform, 235, 441-45.

[5] HermiT OWL reasoner: `http://www.hermit-reasoner.com/`

[6] Example ontology (layers.owl) that demonstrates the application of the pattern: `https://github.com/JoBuRo/layered-structures`

[7] Jansen, L., & Schulz, S. (2014). Crisp islands in vague seas: Cases of determinate parthood relations in biological objects. In Mereology and the Sciences (pp. 163-188). Springer, Cham.

[8] Neal PJ, Shapiro LG, Rosse C. The digital anatomist structural abstraction: a scheme for the spatial description of anatomical entities. Proc AMIA Symp. 1998:423-427