Concept-based versus Realism-based Approach to Represent Neuroimaging Observations

Emna Amdouni[1] and Bernard Gibaud[1,2]
[1]E-health Department, B-com Institute of Research and Technologyy, Rennes, France
[2]LTSI Inserm 1099, Université de Rennes 1, Rennes, France

Keywords: Knowledge Representation, Domain Analysis and Modeling, Concept-based and Realism-based Ontologies.

Abstract: The aim of this paper is to argue why we should adopt a realism-based approach to describe neuroimaging features that are involved in clinical assessments rather than a concept-based approach. This work is a part of a proposal aiming at making explicit the meaning of neuroimaging observations via realism-based ontologies.

1 INTRODUCTION

In most cases, the assessment of radiological findings in clinical practice is still subjective (Rubin et al., 2014) and sometimes error-prone (Rector et al., 1991) (Smith et al., 2006). For instance two radiologists may estimate different volumes of the same tumor or disagree about the presence or not of a lesion in a given brain. For Smith, an automatic production of radiological observations is needed to 'reduce logical contradictions' and to enable advanced imaging research by capturing observations in a standardized format that supports logic-based reasoning. We believe that the use of an ontology could be the appropriate manner to address these challenging points.

However, the semantic description of radiological observations is not a trivial task for two main reasons: first, medical images are semantically rich and they refer to complex entities that may exist or not 'on the side of the patient' at a given period of time (Ceusters et al., 2006). Second, these identified entities should be tracked to follow their evolution through time (Ceusters and Smith, 2005). For example, the nature of David's lesion may change and evolve from benign to malignant at successive time points. This means that in clinical observation statements we will refer to the same entity (David's lesion) but in different ways (absent, malignant, enlarged, etc.) during David's lifetime.

In many research works (Ceusters et al., 2006) (Cimino, 2006) (Smith, 2006) (Smith et al., 2006), authors have highlighted the importance of the distinction between existing entities and non-existing entities on the side of the patient to enable a 'faithful representation' of imaging features. Moreover, they have expressed the need to enable tracking related individual entities over the whole patient's lifetime.

There are two modeling manners that are adopted in literature to semantically describe image contents: a concept-based paradigm (Cimino, 2006) and a realism-based paradigm (Smith, 2006). The concept-based paradigm focuses its modeling on 'concepts', beyond the 'terms' that are used. The realism-based paradigm aligns terms in terminologies on 'existing entities in reality' rather than concepts. The realism-based approach distinguishes between three levels of knowledge, presented in (Smith et al., 2006): 'Level 1: the objects, processes, qualities, states, etc. in reality, Level 2: cognitive representations of this reality on the part of researchers and others, Level 3: concretizations of these cognitive representations in representational artifacts'.

Smith notes that existing medical terminologies define a 'wide variety of universals', but they 'allow direct reference to just a small number of particulars normally just to: human beings, times and places.' Hence, medical terminologies do not refer to concrete existing 'phenomena on the side of the patient', but they only code medical statements in a formal way. As a result, Smith considers that most implementations do not enable 'keeping track of one and the same particular (for example, a specific tumor) over an extended period of time' nor

Amdouni, E. and Gibaud, B.
Concept-based versus Realism-based Approach to Represent Neuroimaging Observations.
DOI: 10.5220/0006084401790185
In *Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2016) - Volume 2: KEOD*, pages 179-185
ISBN: 978-989-758-203-5

'distinguishing between multiple examples of the same particular and multiple particulars of the same general kind'.

The aim of this paper is to argue why a realism-based approach should be adopted to describe neuroimaging features that are involved in clinical assessments rather than a concept-based approach. This work is an initiative towards making the meaning of neuroimaging observations explicit via realism-based ontologies.

2 LITERATURE REVIEW

Concept-based and realism-based methods have been proposed to describe human observations about medical images. These two paradigms propose distinct definitions of the term 'concept'. According to Cimino, the term 'concept' is a 'unit of symbolic processing' in medical terminologies' construction. Smith considers concept-based terminologies as 'collections of elements that may refer or not to concrete entities' (for example, a medical diagnosis concept does not exist in reality but can be modeled via concept-based terminologies) and that 'groups terms by which radiologists express ideas'. Thus, the term 'concept' is conceived by Smith as 'a real world referent of the concept ID that is the class of entities in reality which the concept ID represents'. In the realism-based paradigm, we refer to the real entities themselves as they exist in reality through unique identifiers, whereas in the concept-based paradigm we focus on the representation of data about these entities.

2.1 Concept-based Paradigm

The concept-based paradigm is based on the use of concepts as mind-related entities, i.e. concepts are referred to by terms that are part of a domain-specific lexicon. In (Cimino, 2006), Cimino distinguished between three desiderata that should be respected in the concept-based paradigm: 'non-vagueness', 'non-amibuity' and 'non-redundancy' of concepts. The ultimate objective of this paradigm is to code with an 'ontological view' the representation of concepts in information systems to enable their automatic retrieval. As mentioned by Smith, 'terminologies composed of expressions of ideas lead to difficulties', especially: 1) a non-complete representation of the real world as concepts refer to 'universals' but do not precise to which 'particular instances' of these universals they are referring, 2) a non-adequate modeling of absent entities, 3) a confusing interpretation of data as there is no consistent identification framework of entities, etc.

Visually Accessible Rembrandt Images terminology: The Visually Accessible Rembrandt Images terminology called VASARI terminology (Frederick Nat. Lab for Cancer Research, 2014) is a controlled vocabulary that describes thirty observations of high grade cerebral gliomas (especially glioblastoma multiform or GBM) in conventional Magnetic Resonance Images (MRI). Its main objective consists in standardizing brain tumors' description and facilitating their interpretation by neuro-radiologists. The VASARI terminology was developed by domain experts who have considered the majority of possible MRI brain tumor neuro-radiologists' assessments. Each VASARI feature is represented by a feature number (e.g. F1, F2, etc.) and a set of possible label values to score features. For example, F1 'tumor location' assesses the location of the geographic epicenter and it has six possible label values= {frontal, temporal, insular, parietal, occipital, brainstem, cerebellum.}; F29 and F30 'lesion size' measure the 'largest perpendicular (x-y) cross-sectional diameter of T2 signal abnormality measured on a single axial image only'.

The VASARI terminology is easy to use by neuro-radiologists, especially when user-friendly interfaces are available (e.g., Clear Canvas implementation of VASARI). However, this annotation method adopts a linguistic view rather than an ontological view to generate label values. Thus, these values cannot make explicit to what real entities on the side of the patient the neuro-radiologist is referring in his or her evaluation?

2.2 Realism-based Paradigm

Smith assumes in the realism-based approach that there is 'only one universal objective reality' and that 'only things in reality would be considered'. For example, 'each attribute of the patient is itself a unique entity in reality and it is assigned its own identifier', and as he said 'when a patient's temperature is measured, the measurement is an instantaneous entity, while the polyps seen during a colonoscopy are persisting entities in reality'.

Basic Formal Ontology: The Basic Formal Ontology (BFO) (Grenon et al., 2003) is a realism-based ontology that describes existing entities and relations that exist in reality and are common in all scientific domains. BFO enables a coherent representation of the underlying reality. The use of such realism-based ontologies is recognized as one of the best practices in ontology design for three main reasons. First, it is based on a fundamental separation between entities that persist through the time called 'BFO: continuant' and processes that happen and to which continuants participate called 'BFO: occurrent'. Second, it distinguishes the general from the specific, i.e., what philosophers call 'universal' and 'particular', respectively), and it states that only particulars instantiate universals. Third, its relationships are formally defined in the Relation Ontology (RO) (Smith et al., 2005).

RO is a realism-based ontology that, as BFO does, distinguishes between the universals' level and the particulars' level. Thus, it defines three fundamental types of binary relations listed in (Smith et al., 2006): <universal, universal>, e.g.: hand **part_of** body, <particular, particular>, e.g.: David's hand **part_of** David and <particular, universal>, e.g.: David's hand **instance_of** hand.

Referent tracking: Referent tracking (RT) (Ceusters and Smith, 2005) is a paradigm for data entry and retrieval that identifies and directly refers to relevant 'concrete individual entities' that are fundamental for the description of the clinical context of a specific patient. Its objective is to 'reduce ambiguous reference within Electronic Health Records (EHR)'. RT refers explicitly to those existing entities through the assignment of unique identifiers. For example, we can assign a unique identifier to a tumor of a specific patient. Thus, it becomes possible to perform multiple measurements on the same tumor (to check for consistency), to perform different measurements on the same tumor (to correlate findings and to report multiple results on the same patient). Identifiers can be adopted even in the representation of negative clinical findings to refer to non-existing entities 'on the side of the patient' (Ceusters et al., 2006).

Realism-based approach versus concept-based approach: Unlike the concept-based approach the realism-based approach refers to only real entities as they exist in reality. In his conceptual view, Cimino does not address 'what it is on the side of the patient'. Therefore, the realism-based approach insures a faithful representation of a portion of reality by explicitly referring to instances of universals and representing interrelations between them. Unique identifiers, in the realism-based approach, are attributed to instances of universals rather than concepts. Hence, the identity of the same particular can be preserved at successive time points. In contrast, in the concept-based, approach when the entity changes its aspect, a new code is assigned to the referred entity to expresss this evolution. Thus, we can not generate an history about a specific entity's evolution. This limit of the concept-based approach makes it impossible to follow entities along their evolution (entities that no longer exist or change of type, etc.).

However, in terms of implementation, the concept-based approach is a simple data annotation solution given that it does not necessitate the development of particular systems to generate entities' identifiers and to handle complex entities.

2.3 Standardized Formats for Recording Imaging Observations

Annotation and image markup model: The Annotation and Imaging Markup (AIM) model (Rubin et al., 2008) is an information model that can be stocked as an XML-based file format to describe the minimal information necessary to record an image annotation. The AIM: semantic image group distinguishes between two elementary classes related to the medical image content: AIM: imaging physical entity that denotes a referent (for example the mass on the side of the patient) and AIM:image observation entity that represents references (for example the mass observed on the medical image). The AIM model has defined two distinct qualities, accordingly: the first one describes real entities (for example: enlarged or not) and the second one describes the appearance of physical objects on the medical image. The AIM model has been used to track lesion measurements across imaging series in clinical trials (ePAD (Rubin et al., 2014)), lesions were recorded as AIM image annotations and tagged by a unique identifier.

To summarize, the AIM model has introduced the most relevant entities in image annotation, but its implementation lacks formal semantics, since it is not based on ontologies. As a consquence, we cannot represent complex entities or perform logic-based reasoning to infer new knowledge about image content.

DICOM SR: The DICOM standard (Digital Imaging and Communications in Medicine) specifies a data structure for structured reports (Clunie, 2007) as a set of rules constraining their organization and a vocabulary specifying which codes should be used and the associated code meanings covering the domain of imaging observations. DICOM SR enables the representation of radiological observations. It includes measurements and qualitative assessements, their relationships with image evidence and with the clinical interpretation of the clinician. However, DICOM SR suffers of several limitations. The use of standard terminology is strictly limited to the use of standard codes borrowed from several external terminology resources (mainly from SNOMED CT). The way these terms are modeled in the SNOMED CT ontology (e.g. subsumption links) is not exploited. Moreover, the relations between terms that are used in the SR model is not based on those existing in, e.g. SNOMED CT, but are specifically defined in DICOM SR. Finally, no query language exist to retrieve information from an SR tree, thus requiring to export a SR tree content to some relational database or XML data structure (e.g. AIM serialization) to perform any queries on content.

3 WHY SHOULD A REALISM-BASED APPROACH BE ADOPTED TO REPRESENT NEUROIMAGING OBSERVATIONS?

Cerebral tumor assessment consists in the characterization of the anatomical, functional and molecular aspect of the tumor. These assessments are about basic brain tumor entities on the side of the patient. For this reason, the recording of neuroimaging observations should allow radiologists to make assertions about these basic entities in the neuroimaging domain and track their evolution.

The coverage of all neuroimaging information involved in cerebral tumor assessment is impossible since no consensual source exist to specify precise requirements of this domain. To address this difficulty, we have limited our study to the domain covered by the VASARI terminology. Actually, VASARI constitutes a representative use case, i.e. raising the most typical situations that need to be modeled and calling for relevant solutions to some important challenges related to standard web languages' constraints and restrictions. Our proposed semantic modeling of VASARI features is made in OWL DL (Ontology Web Language) and is based on the realist ontology Basic Formal Ontology (BFO).

During a brain tumor assessment, the neuro-radiologist assigns a value to each VASARI imaging feature, thus providing a standardized description of relevant aspects of the tumor that should be taken into consideration in clinical decision making. The labelling of these imaging features involves different kinds of entities: physical parts, qualities related to physical objects and volume and size measurements. According to the VASARI terminology, the basic physical entities that characterize some brain tissue abnormality are: the cerebral tumor, the cerebral tumor geographic epicenter, cerebral tumor components (namely: contrast enhanced region, non contrast enhanced region, necrotic part, edematous component and cerebral tumor margin) and the outside of the margin of a cerebral tumor or a part of a cerebral tumor.

The current formalization of neuroimaging information based on the VASARI terminology ensures a comprehensive and a simple description of cerebral tumors contained in medical images by a simple labelling of images. However, the VASARI terminology provides only a free text definition of the meaning of VASARI scores e.g., no formal axioms are defined to formalize such meaning in a logical language for example what is the quality that is measured? What individual entities are involved in the evaluation of an imaging criterion? Thus, the neuroimaging features as currently presented with the VASARI terminology are not instantiable and do not describe the real 'phenomena on the side of the patient'.

In our semantic model, we have described the thirty VASARI features. However, only some of them will be cited as illustrative examples in this paper. Our methodology to design a realism-based ontology is composed of five main steps that are summarized as follows: First, find for each VASARI feature F the meaning of the studied aspect and sort its possible configurations. The latter arise from the list of possible values allowed for each criterion. Second, identify and describe the key entities involved, bearing in mind the concern to primarily focus on real entities. Third, relate entities to existing ontologies, or create new classes by specializing existing ones. Fourth, specify the axioms characterizing these entities. Finally, make sure that all possible configurations for each feature F can be modeled in a formal way.

3.1 What Particulars Are Referred to in Each Imaging Feature?

The VASARI feature F3 evaluates if 'the geographic center or the enhancing component involves the eloquent cortex (motor, language, vision) or key underlying white matter?' For example: F3=2 means that the epicenter of the given cerebral tumor has affected the Broca's area. As it is formulated, it is not explicit if the attribution of multiple values is allowed or not. Besides, we cannot determine anatomical regions that are affected by the cerebral tumor.

The VASARI feature F15 entitled 'edema crosses midline' evaluates if the cerebral edema component 'spans white matter commissures extending into contralateral hemisphere'. A formal representation of this feature requires a direct reference to specific cerebral edema components that are located in distinct cerebral hemispheres and that are adjacent to some white matter commissure.

3.2 Representation of Negative Neuroimaging Observations

In neuroimaging terminologies, non existing entities are expressed with negative qualifiers and expressions: 'none', 'no', 'indeterminate', 'does not', 'not applicable', etc. These expressions do not refer to anything in reality. In the VASARI terminology, this negation can concern two distinct categories of continuants: independent continuant and dependent continuant. Based on this classification we have distinguished two modeling cases:

Case 1 where an independent continuant is totally absent or is not located in a given region of the patient's brain. Here, we refer to assessments that express for example that a cerebral tumor does not have as part an enhancing cerebral tumor component or is not located in the cerebral brain cortex of the patient.

Case 2 where a dependent continuant is absent: this case comprises two categories of statements: statements that refer to absent qualities and those that express the lack of a disposition for a given independent continuant, e.g., cerebral tumor. To model these two subcases we have used a simple logical negation to define entities that do not reflect anything in reality. For example, the formal representation of a non contrast region of the tumor can be defined as follows: 'Non enhancing cerebral tumor component'≡ **is_a** 'Cerebral tumor component' and not (**has_disposition** some 'Disposition to be enhancing').

3.3 Representation of Spatial Knowledge

Cerebral tumors may change their location during their existence and occupy different spatial regions. Thus, to ensure a correct evaluation of tumor evolution we need to formally represent how the cerebral tumor and its components are situated in space at different periods of time. In our semantic model, we have modeled spatial knowledge (Brandon et al., 2013) inside and outside the tumor or the anatomical structure.

Inside: Two types of spatial inclusions are mentioned, in the VASARI terminology: containment (non tangential part) and overlapping (tangential proper part). Containment is denoted by these natural language expressions like 'within', 'portion of', 'comprise of' whereas overlapping is denoted by the term 'invasion'. Formally, these relationships are represented by **part_of** and **located_in**, respectively.

Outside: Here, we express the proximity of a given entity to another one. In the VASARI terminology, there are two types of proximity: adjacency expressed with terms such as 'surrounding' and 'adjacency', and separation denoted by terms like 'not contiguous' and 'separated'. We have reused the RO relation **adjacent_to** to express that an entity is near another entity and the class BFO: relation quality to represent the two qualities 'connected to another cerebral tumor component' and 'contiguous with cerebral tumor'.

3.4 Representation of Complex Entities

The representation of the extent of the resection of a given cerebral tumor component (enhancing, non enhancing or edematous part) is not simple given that it evaluates the volume change of a given cerebral tumor component at two different time points (before and after the surgical intervention). To model this type of features (namely F26, F27 and F28), there are two modeling manners:

Proposition 1: We consider that the cerebral tumor component will not preserve its identity after and before the surgery. Thus, we will identify two distinct entities: cerebral tumor component before surgery and cerebral tumor component after surgery. In this case, the measured quality is the same, i.e. the volume, but measured volume values are distinct.

Proposition 2: We consider that we will refer to the same cerebral tumor component instance at two distinct time points. So, this instance will have two distinct volume qualities, namely volume before surgery and volume after surgery.

The interpretation of extreme values of volume ratio will be as follows: 0% means we have not resected any part of the cerebral tumor component. Thus, the 'measured volume value' before the surgery is the 'measured volume value' after the surgery (in P1) or the 'quality volume' before the surgery is the 'quality volume' after the surgery (in

P2). 100% (in P1 and P2) means that the cerebral tumor component is totally resected and thus the instance of cerebral tumor component no longer exists.

4 DISCUSSION

Our study of the VASARI terminology shows that a concept-based approach cannot insure a faithful representation of neuroimaging observation and that a realism-based orientation should be adopted in this context. We can list three main challenging points that can be avoided when such an approach is adopted: 1) a formal and an explicit description of imaging features' meanings, 2) a distinction between existing and non existing entities and 3) a following of entities' evolution.

In this paper, we have first discussed the advantages and limits of concept-based and realism-based approaches, second we have outlined the procedure that should be followed to track relevant entities in the domain of neuroimaging and finally we have explained how the realism-based orientation could be used to answer to the domain's requirements?

We think that the adoption of a realistic view can help automating the generation of neuroimaging assessments, via image processing techniques, and covering important domain's needs (tracking particulars over the whole patient's lifetime, detecting absent entities, representing complex situations, etc.). The transformation of neuroimaging labels into quantitative and qualitative information will: reduce ambiguity in clinical statements, improve the reproducibility of assessments in computer assisted detection and enable semantic reasoning about involved particulars.

The main limit of our proposal is that we have not addressed the problem of variability between annotators. This problem of logical contradiction, studied in (Rector et al., 1991), is due to the fact that radiologists describe what they observe based on their thoughts and experiences, as consequence they may describe differently identified entities (for example, cerebral tumors) and produce different assertions about these entities. In this case, a problem of logical conflicts may occur, for example, David's cerebral tumor **contains** an enhancing cerebral tumor component and David's cerebral tumor **does not contain** an enhancing cerebral tumor component. As underlined in (Smith et al., 2006), this problem of logical contradiction is not handled in concept-based nor in realism-based approaches.

The second limit of our proposal, implemented in OWL, is that it does not generate temporalized instances with this logic-based language (Smith et al., 2006). We think that taking into consideration the temporal aspect in the representation of neuroimaging features is needed especially in longitudinal imaging studies to, for example, evaluate cancer treatment response. In this context, we recommend to select a logic-based language that is capable to represent ternary relationships.

To conclude, the management of information about imaging features (measurement values, qualities, lesion components, lesion localization, etc.) will support clinical research on the discovery and the validation of new imaging biomarkers. Moreover, such information may also be used for clinical decision support, for example, predicting patient survival based on GBM features.

REFERENCES

Brandon, B., Vinay, C., Nikhil, D., 2013. A Vocabulary of Topological and Containment Relations for a Practical Biological Ontology. Spat. Inf. Theory, Lecture Notes in Computer Science 418–437.

Ceusters, W., Elkin, P., Smith, B., 2006. Referent tracking: the problem of negative findings. Stud. Health Technol. Inform. 124, 741–746.

Ceusters, W., Smith, B., 2005. Tracking referents in electronic health records. Stud. Health Technol. Inform. 116, 71–76.

Cimino, J.J., 2006. In defense of the Desiderata. J. Biomed. Inform. 39, 299–306..

Clunie, D.A., 2007. DICOM Structured Reporting and Cancer Clinical Trials Results. Cancer Inform. 4, 33–56.

Frederick Nat. Lab for Cancer Research, 2014. VASARI Research Project - The Cancer Imaging Archive (TCIA) [WWW Document]. URL https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project (accessed 7.14.16).

Grenon, P., Smith, B., Goldberg, L., 2003. Biodynamic ontology: applying BFO in the biomedical domain. Stud. Health Technol. Inform. 102, 20–38.

Rector, A., Nowlan, W., Kay, S., 1991. Foundations for an electronic medical record. Methods Inf. Med. 30, 179–186.

Rubin, D.L., Willrett, D., O'Connor, M.J., Hage, C., Kurtz, C., Moreira, D.A., 2014. Automated Tracking of Quantitative Assessments of Tumor Burden in Clinical Trials. Transl. Oncol. 7, 23–35.

Rubin, D., Mongkolwat, P., Channin, D., 2008. A semantic image annotation model to enable integrative translational research. Summit Transl. Bioinforma. 2009, 106–110.

Smith, B., 2006. From concepts to clinical reality: an essay on the benchmarking of biomedical terminologies. J. Biomed. Inform. 39, 288–298.

Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A.L., Rosse, C., 2005. Relations in biomedical ontologies. Genome Biol. 6, R46.

Smith, B., Kusnierczyk, W., Schober, D., Ceusters, W., 2006. Towards a Reference Terminology for Ontology Research and Development in the Biomedical Domain.