Using semantic dependencies for consistency management of an ontology of brain-cortex anatomy.
Olivier Dameron, Mark Musen, Bernard Gibaud

To cite this version:

HAL Id: inserm-00152177
https://www.hal.inserm.fr/inserm-00152177
Submitted on 29 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Using semantic dependencies for consistency management of an ontology of brain–cortex anatomy

Olivier Dameron\textsuperscript{a,b,\*}, Mark A. Musen\textsuperscript{b}, Bernard Gibaud\textsuperscript{c}

\textsuperscript{a}EA 3888, Université de Rennes 1, 2 Rue Henri Le Guilloux, F-35033 Rennes, France
\textsuperscript{b}Stanford Medical Informatics, Stanford University, 251 Campus Drive, X-215, Stanford, CA 94305, USA
\textsuperscript{c}Équipe/Projet VisAGeS U746, INSERM/INRIA/CNRS/Université de Rennes 1, 2 Avenue du Pr Léon Bernard, F-35043 Rennes Cedex, France

Received 28 September 2004; received in revised form 19 September 2006; accepted 22 September 2006

KEYWORDS
Ontology consistency; Ontology maintenance; Dependencies between concepts or relationships

Summary

Objective: In the context of the Semantic Web, ontologies have to be usable by software agents as well as by humans. Therefore, they must meet explicit representation and consistency requirements. This article describes a method for managing the semantic consistency of an ontology of brain–cortex anatomy.

Method: The methodology relies on the explicit identification of the relationship properties and of the dependencies that might exist among concepts or relationships. These dependencies have to be respected for insuring the semantic consistency of the model. We propose a method for automatically generating all the dependent items. As a consequence, knowledge base updates are easier and safer.

Result: Our approach is composed of three main steps: (1) providing a realistic representation, (2) ensuring the intrinsic consistency of the model and (3) checking its incremental consistency. The corner stone of ontological modeling lies in the expressiveness of the model and in the sound principles that structure it. This part defines the ideal possibilities of the ontology and is called realism of representation. Regardless of how well a model represents reality, the intrinsic consistency of a model corresponds to its lack of contradiction. This step is particularly important as soon as dependencies between relationships or concepts have to be fulfilled. Eventually, the incremental consistency encompasses the respect of the two previous criteria during the successive updates of the ontology.

\* Corresponding author at: EA 3888, Université de Rennes 1, 2 Rue Henri Le Guilloux, F-35033 Rennes, France. Tel: +33 2 99 28 92 45; fax: +33 2 99 28 41 60.
E-mail address: olivier.dameron@univ-rennes1.fr (O. Dameron).

0933-3657/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.artmed.2006.09.004

1. Context

A symbolic model allows developers to represent general knowledge about a domain and the meaning that is commonly associated with it. This knowledge can be used by itself (e.g., for teaching), or indirectly as a reference to process specific facts (e.g., to assist queries or data retrieval). In the latter case, symbolic models are perceived as a key feature to provide software assistance for tasks that now require domain-aware intervention by a human.

As interoperability of these software applications is desirable, shared conceptual models, and specifically ontologies, play a major role in a Semantic Web context [1]. Since these models are to be usable by software, they must meet explicit representation and consistency requirements.

For medical applications, anatomy provides a common reference used to reason about pathology or localization of functional activity [2, 3]. The Foundational Model of Anatomy (FMA) [4] and Galen [5] are two major conceptual models that provide a symbolic representation of human anatomy. However, neither of them provides a satisfactory representation of brain—cortex anatomy. The major sources of neuro-anatomical knowledge are paper-based atlases [6, 7] and terminological systems such as Neuronames [8].

1.1. Existing model

We are working on an ontology of brain—cortex anatomy. Our goal is more to formalize existing knowledge than it is to propose new anatomical concepts or relationships. Our model has been described in previous publications [9, 10]. It comprises 304 concepts and 1254 relationships that represent the organization of anatomical structures. Because the brain surface presents complicated folding patterns, typical anatomical structures are gyri (the bulges of cerebral matter, similar to hills), sulci (the hollow foldings, similar to valleys) and lobes (sets of gyri).

Our model’s taxonomy hierarchy is composed of three levels. First, the generic level contains concepts such as Lobe or Sulcus, and is mainly used to define the domain and range of the relationships. Second, the abstract level represents a prototypical brain hemisphere, and contains concepts such as Frontal Lobe or Central Sulcus. Third, the lateralized level is used to represent left/right asymmetries, and contains concepts such as Left Frontal lobe.

For mereology, the model identifies several relationships such as hasDirectAnatomicalPart, hasAnatomicalPart, hasSegment and their properties, inspired from previous theoretical works [11, 12].

The model also represents neighborhood relationships such as the separation of two cortical structures by a sulcus, anatomical continuity, and sulci connection.

1.2. Checking the model’s consistency

In this context, managing the semantic consistency of the ontology has been one of our main concerns. This work has consisted in checking that the model reflects reality, and that the relationship properties are respected. The last point has led us to identify dependencies among relationships. This article describes some of these dependencies and proposes an original method for making sure that they are respected during successive updates. This method consists of automatically generating all the dependent relationships, which also make the knowledge base maintenance easier.

The “Realism of representation” section describes our efforts to keep our model as close as possible to reality. The “Intrinsic consistency” section describes the identification of dependencies among relationships, and their representation by implication rules which can be used to generate a self-consistency of a version of the ontology. The “Incremental consistency” section describes how to make sure that the successive updates are all self-consistent and provide the expected modifications.

2. Realism of representation

The adequacy of the model with respect to some reality is a core aspect of ontological modeling. It ensures that the definitions and the propositions derived from the model are acceptable. Since reality is hard to define and can be relative, canonical knowledge [4](i.e., derived from generalization and synthesis of previous observations) provides at least a gold standard [13]. For our work on brain anatomy, reference atlases [6, 7] and discussion with an
expert provided the base of the canonical knowledge.

The correspondence between the concepts of a symbolic model and the elements of some reality is achieved through an interpretation function [14], which maps every concept to its individual instances. The structure of the model defines the possible interpretation functions. A lax model would allow interpretation functions that associate concepts that do not match generally admitted knowledge to concrete situations. Conversely, a restrictive model would dismiss the interpretation functions that associate desired concepts to concrete situations. This section describes how we tried to make our symbolic model as restrictive as possible with regard to anatomical variability.

2.1. All reality

An ontology is a simplified view of some reality. However, an ontology has to comply with all the situations of the domain of study. For instance, our model of brain anatomy has to cope with a precentral sulcus composed of two segments for one individual, as well as a precentral sulcus composed of four segments for another one.

For our brain anatomy ontology, the main difficulties lay in left/right asymmetries between the two hemispheres, as well as in inter-individual variability. The acknowledgment of this variability and its explicit representation in our model is particularly apparent in part/whole as well as in topological relationships, where a distinction has to be made between mandatory and possible relationships. Necessary conditions are represented by the existential operator (\(\exists\)). Possible conditions are represented by the universal operator (\(\forall\)). For instance, "the precentral sulcus (precs) must have a superior segment (sup-precs) and an inferior segment (inf-precs), and can have an intermediate (int-precs) and a marginal segment (margin-precs)" is represented by "all the segments of precs are sup-precs or inf-precs or int-precs or margin-precs; and there is a sup-precs; and there exists an inf-precs". In addition, existence probabilities for concepts as well as for relationships are specified whenever possible.

Modeling all reality is pretty easy by reducing the constraints. Therefore, lax models are favored here.

2.2. Only reality

Ideally, an ontology must not allow developers to describe things other than those in the reality being modeled. A model of anatomy that would allow a brain hemisphere to have any number of lobes, or two frontal lobes, cannot be considered as a good model. Therefore, the model has to enforce enough constraints in order to reject any bad interpretation of the reality. We took this point into account for specialization, composition and topological relationships.

In the taxonomic hierarchy, the distinction between the generic, abstract, and lateralized levels, as well as the consideration that the concepts of a same level are mutually exclusive (e.g., a lobe cannot be both a frontal lobe and a parietal lobe) conform to this principle.

For mereological relationships, both the cardinality constraints and the partitioning principle that requires that anatomical structures have no common part also play important roles. For instance, we do not simply state that "a hemisphere is composed of five lobes; frontal lobe is a lobe; parietal lobe is a lobe; temporal lobe is a lobe, occipital lobe is a lobe and limbic lobe is a lobe", as most symbolic models of anatomy would do. We stated that a hemisphere has five direct anatomical parts that include exactly one frontal lobe, exactly one parietal lobe, etc.; these five lobes are mereologically mutually disjoint.

For topological relationships, representation using binary relationships that a sulcus separates two cortical structures, just as a river separates two regions, could lead to erroneous inferences. Fig. 1 illustrates such situations. If we use a binary relationship to represent that a sulcus \(S\) is a boundary of a cortical structure (e.g., \(G_1\)) as shown in the middle column, then we are unable to infer correctly that \(S\) separates \(G_1\) from \(G_3\) but not from \(G_2\). The bottom of Fig. 1 shows another typical situation where some erroneous separations cannot be ruled out. Therefore, we had to use a ternary \(\text{separates}\) relationships (right column of Fig. 1).

3. Intrinsic consistency

There are important dependencies among the relationships in our model of brain anatomy. The various dependencies we could identify are described in Section 3.1. These dependencies can be seen as consequences of the properties of the relationships.

These dependencies could be modeled by implication rules. Examples of such rules are provided in Section 3.2.

3.1. Dependencies between relationships

3.1.1. Specialization dependencies

Specialization-related dependencies occur between a general concept and a more specific one. Such
dependencies are similar to those of inheritance for object-oriented models. Although very simple, such dependencies still have to be taken into account.

In our model of anatomy, specialization dependencies occur between the three taxonomic levels. For instance, a Sulcus (generic level) is filled with cerebro-spinal fluid. Therefore, the Central Sulcus (abstract level) which is subsumed by Sulcus, is also filled with cerebro-spinal fluid, and so are the Left Central Sulcus and the Right Central Sulcus (lateralized level).

Specialization dependencies can also take place between relationships. Thus, the existence of the hasDirectAnatomicalPart relationship between two anatomical structures implies that they are also linked by the broader hasAnatomicalPart relationship. Similarly, if a Sulcus isBranchOf another one, both of them also have to be linked by the isConnectedTo relationship.

3.1.2. Dependencies between mereological relationships
The dependencies between part/whole relationships are mainly consequences of the taxonomy of mereological relationships (Fig. 2) and of the transitive property of some of them. Particularly, the isDirect... relationships are non-transitive sub-relationships of transitive ones. This is a standard practice both in programming and in knowledge representation. For example, the Orbital Pars of Inferior Frontal Gyrus isDirectAnatomicalPartOf Inferior Frontal Gyrus. The relation isDirectAnatomicalPartOf is a sub-relation of isAnatomicalPartOf. Therefore, the latter also holds between the two cortical structures. Similarly, Inferior Frontal Gyrus isDirectAnatomicalPartOf Frontal Lobe. It follows that Inferior Frontal Gyrus isAnatomicalPartOf Frontal Lobe. As the isAnatomicalPartOf relationship is
considered to be transitive (whereas isDirectAnatomicalPartOf is not), it also must hold between Orbital Pars of Inferior Frontal Gyrus and Frontal Lobe.

The spatial extensions of anatomical structures constitute another example of dependencies (Fig. 3).

Indeed, there is a mereological hierarchy between the spatial extensions of an anatomical structure (Fig. 4). This hierarchy combines with the mereological hierarchy of anatomical structures, as mereological relationships between anatomical structures implies mereological relationships between their spatial extensions (Fig. 5). For instance, the VisibleCorticalZone of a cortical anatomical structure isSubAreaOf the ExtendedCorticalZone of the same structure. This is true for the PreCentral Gyrus as well as for the Frontal Lobe. But since the former isAnatomicalPartOf the latter, the VisibleCorticalZone (respectively ExtendedCorticalZone) of PreCentral Gyrus isSubAreaOf the VisibleCorticalZone (respectively ExtendedCorticalZone) of Frontal Lobe. This example shows that dependencies can occur between relationships such as isVisiblePartOf and isAnatomicalPartOf that are not sub-relations of each other.

3.1.3. Dependencies between topological relationships

The dependencies between topological relationships are mainly due to the taxonomy of these relationships. For instance, if a sulcus separates two cortical structures, then these structures also have to be contiguous.

The duality between the configuration of the sulci and that of the gyri is another example of dependencies. However, these dependencies are hard to model and have not yet been taken into consideration.

Figure 3 Possible spatial extensions of a gyrus. The cerebral zone is referred to in surgical procedure. The extended cortical zone is referred to in functional activity studies. The visible cortical zone is referred to when teaching neuroanatomy.

Figure 4 Mereological relationships among the spatial extensions of an anatomical structure.

Figure 5 Dependencies between the mereological hierarchy of anatomical structures and that of their spatial extensions. The compound is on the diamond side of mereological relationships.
3.1.4. Combined dependencies

Of course, it is also possible to combine the three previous kinds of dependencies which makes it harder to categorize them. These combinations are particularly interesting because they involve dependency patterns that are more complex than simple sub-relations. For instance, if two cortical structures are separated by a segment of a sulcus, they are also separated by this sulcus (e.g., Pre-Central Gyrus and Superior Frontal Gyrus are separated by Superior PreCentral Sulcus; therefore, they are also separated by PreCentral Sulcus), thus combining mereological and topological dependencies.

Another example of dependency combining both mereological and topological relationships also involves relationship properties. Two cortical structures are anatomically continuous if and only if their visible parts are externally connected. If one of the two cortical structures is an anatomical part of a whole, but the other is not a part of this whole, then the visible part of the whole can be proved to be externally connected to the visible part of the second anatomical structure. Fig. 6 a shows a schema of such a dependency. For example, Pre-Central Gyrus and Opercular pars of Inferior Frontal Gyrus are anatomically continuous. Since the Opercular pars is an anatomical part of Inferior Frontal Gyrus and since Inferior Frontal Gyrus and PreCentral Gyrus are mereologically disjoint (they do not have any common part), they also have to be anatomically continuous. In addition, this inferred relationship can be used iteratively to apply the same principle (which is equivalent to using the transitive property of isAnatomicalPartOf). Fig. 6 b–d illustrates the successive application of this principle.

This approach can be extended to anatomical contiguity or the separation of two cortical structures by a sulcus. Thus, the fact that Central Sulcus separates Frontal Lobe and Parietal Lobe can be seen as a consequence of the fact that Central Sulcus separates PreCentral Gyrus (a part of Frontal Lobe) and PostCentral Gyrus (a part of Parietal Lobe).

Finally, specializing abstract level concepts into lateralized concepts also generates dependencies.

3.2. Consistency rules

The previous dependencies can be represented as implication rules. Such rules, along with the relationship properties, constitute knowledge about anatomical knowledge. They belong to a level separate from that concepts and relationships.

The implications can form the basis of an inference engine that automatically generates all the dependent concepts and relationships.

3.2.1. Abstract restricted model

We only maintain an abstract restricted model composed of:

- the concepts of the abstract level (i.e., non lateralized, such as Central Sulcus and Frontal Lobe);
- all the independent relationships;
- a restricted base of asymmetry-specific facts, such as the different existence probabilities for the left and right intermediate precentral sulcus.

Typically, it consists in representing taxonomic relationships, direct mereological relationships, and topological relationships among the smallest parts.

3.2.2. Extended abstract model

The extended abstract model is generated automatically. This step consists of inferring all the dependent relationships among composed anatomical structures. 59.7% of the relationships from the extended abstract model are automatically created [10].

Figure 6 Relationships depending on hasAnatomicalPart (the compound is on the diamond-side of the line) and on topological relationships of continuity, contiguity or separation by a sulcus (in the latter case, only the two cortical structures are represented on the schema). Example: for $a$, $G_1=$ Opercular Pars of Inferior Frontal Gyrus, $G_2=$ PreCentral Gyrus, $G_3=$ Inferior Frontal Gyrus. The separation of $G_3$ and $G_2$ depends on that of $G_1$ and $G_2$ by PreCentral Sulcus.
3.2.3. Extended lateralized model
The extended lateralized model is generated by applying specialization rules for lateralization to the extended abstract model. These rules:

- create the lateralized concepts as subconcepts of those defined on the abstract level (e.g., Left Frontal Lobe and Right Frontal Lobe are subsumed by Frontal lobe);
- add consistency statements (e.g., Left Frontal Lobe and Right Frontal Lobe are taxonomically disjoint, and Frontal Lobe is equivalent to Left Frontal Lobe or Right Frontal Lobe);
- generate all the required relationships (e.g., from the statement "Frontal Lobe hasAnatomicalPart PreCentral Gyrus", we would infer that Frontal Lobe hasAnatomicalPart Left PreCentral Gyrus (respectively, right) and that Left Frontal Lobe (respectively, right) hasAnatomicalPart Left PreCentral Gyrus (respectively, right)).

4. Incremental consistency

Managing incremental consistency consists in making sure that intrinsic consistency is still respected after an update of the knowledge base, and that the result meets the "realism of representation" requirement. It can be reduced to answering the following questions:

1. Does every concept and relationship that we wanted to add belong to the model? For instance, if we add a part for a gyrus, we want this structure to be a part of every anatomical concept the gyrus is a part of.
2. Did the consistency rules generate any concepts or relationships that do not correspond with anything in canonical knowledge? For instance, a wrong inference rule will generate erroneous relationships.
3. Has every concept and relationship that we wanted to remove actually disappeared?
4. Did we remove from the model more than we should have? For instance, removing a relationship in order to fix the model has for consequence of removing all the dependent relationships, some of which being right.

Because this step consists in comparing the result with canonical knowledge, it has to be performed manually by a (human) domain expert. However, a simple tool has been developed to assist this task.

Every update of the knowledge base only takes place in the restricted abstract model. The abstract and lateralized extended models are then regenerated automatically. A simple XML Style-sheet helps the domain specialist to compare them with their previous versions. As a result, an HTML page is generated which explicitly represents in green all the concepts and relationships that have been added, and in red those which have been removed, similar to the diff command.

5. Discussion

As we are confronted with an increasing number of concepts and relationships, maintaining the ontology's consistency becomes more and more difficult. In addition, the growth of the model is complicated by the need to add a lot of integrity constraints to the model so that it is not too lax. Therefore, our approach seems to be more and more relevant.

Unfortunately, to our knowledge, none of the main symbolic models of anatomy such as the Digital Anatomist Foundational Model and Galen supports an explicit representation of the dependencies among concepts or relationships. This point is particularly important, since both of these ontologies have to handle concepts and relationships that number in the tens of thousands. Galen's sanctioning statements [15] are assertions preventing impossible situations (e.g., "fracture of the eyelid") or redundant ones (e.g., "the hand which is a part of the arm"). They play an important role in Galen consistency, but do not address semantic dependencies between relationships.

Although our approach has only been applied to a model of the brain cortex, it seems that the principle could be extended to any anatomical model. Moreover, it could also be extended to other domains. However, anatomical knowledge is rather stable. Other domains such as pathology or the study of brain functions are more likely to evolve, which would require in addition a management of obsolescence—something we have not studied.

In addition to being used for maintaining a standalone ontology, semantic dependencies can also be used as additional clues for automatically reconciling several ontologies of a same domain or for combining ontologies of different domains. For example, Reconciling two ontologies of one domain now relies on syntactic similarities between the names of the concepts and/or on similarities of their relationships. Semantic dependencies
between concepts or relationships provide additional constraints. Moreover, semantic dependencies are also of particular importance when establishing mappings between domains. For instance, pathology located in a part of an anatomical structure may also need to be recognized as located in the anatomical structure overall. Schulz provides an interesting analysis of this kind of problems [16,17]. These capabilities are needed considering the role of anatomy as a localization reference, and its use in application contexts that require automatic reasoning.

The dependencies identified in this article, and their usage to maintain semantic consistency of an anatomic model are beyond the scope of logical consistency-checking tools such as ConsVISor [13] or FaCT². For instance, ConsVISor would not issue any warning if the central sulcus separates the precentral and postcentral gyri but not the frontal and parietal lobes. This paper describes the management of consistency from the modeling point of view. It does not rely on any representation formalism. However, it turned out that the consistency rules could not be easily represented in ontology languages such as OWL [18]. Extensions such as SWRL³ could provide very interesting future extensions. They would allow to represent explicitly some consistency constraints to map anatomy to pathology (e.g., to express that a tumor located in a part of an organ has also to be considered as a tumor located in the organ itself).

The functionality of the script used for managing the incremental consistency is similar to that of the diff command or of the PROMPT plugin for Protégé [19] (but less flexible). However, the usage of a specific modeling environment is beyond the scope of this article.

By automatically generating more than 59% of the relationships, our approach makes the task of the curator easier, less error-prone and hopefully less tedious. However, choosing the appropriate modifications in the abstract restricted model requires a good understanding of the existing dependency rules. Here again, it is possible to devise some tools for assisting the curator and detecting any principle violation. Eventually, if any problem is detected by the domain expert during the enforcing of incremental consistency, the curator will be in charge of determining if it comes from a modeling error or from an erroneous rule.

6. Conclusion

Our effort to identify explicitly the properties of the relationships we used, as well as the experience of building the ontology, allowed us to identify dependencies among concepts and relationships. The explicit representation of these dependencies is important for the representation of the semantics of the domain. In addition, it turned out that it can be helpfully used to assist in the management of the knowledge base and to ensure the model’s semantic consistency.

The method we adopted consists in maintaining only a core set of independent concepts and relationships. All the dependent items are then automatically generated. A domain expert still have to manually screen the result in order to make sure that it is correct with regard to canonical knowledge.

Acknowledgements

Xavier Morandi assessed the knowledge base and shared his experience of neuroanatomy. Christine Golbreich provided numerous insights and constructive remarks for rules modeling.

References

Using semantic dependencies for consistency management of an ontology


