Using C-OWL for the Alignment and Merging of Medical Ontologies

Heiner Stuckenschmidt[1], Frank van Harmelen[1]
Paolo Bouquet[2,3], Fausto Giunchiglia[2,3], Luciano Serafini[3]

[1] Vrije Universiteit Amsterdam
[2] ITC-IRST, Trento
[3] DIT - University of Trento

January 28, 2004

Abstract

A number of sophisticated medical ontologies have been created over the past years. With their development the need for supporting the alignment of different ontologies is gaining importance. We proposed C-OWL, an extension of the Web Ontology Language OWL that supports alignment mappings between different, possibly incompatible ontologies on a semantic level. In this paper we report experiences from using C-OWL for the alignment of medical ontologies. We briefly review key concepts of the C-OWL semantics, explain the setting of the case study including some examples from the alignment and discuss the possibility of reasoning about the mapping based on the C-OWL semantics We conclude by arguing that C-OWL provides an adequate framework for aligning complex ontologies in the medical domain.

Keywords: Biomedical Knowledge representation, validation and maintenance; Knowledge Representation Languages; Terminology Integration

1 Introduction

The need for terminology integration has been widely recognized in the medical area leading to a number of efforts for defining standardized terminologies. It is, however, also acknowledged by the literature, that the creation of a single universal terminology for the medical domain is neither possible nor beneficial, because different tasks and viewpoints require different, often incompatible conceptual choices [Gangemi et al., 1998]. As a result a number of communities of practice have been evolved that commit to one of the proposed standards. This situation demands for a weak notion of integration, also referred to as alignment in order to be able to exchange information between the different communities.

In [Bouquet et al., 2003] we argued that the current design of the web ontology and its semantics is not suitable for situations where different view on the same domain have to be aligned in a loose way. We proposed an extension of the OWL semantics that allows the specification of semantic relations between different OWL models. The resulting notion of contextualized ontologies can provide such an alignment by allowing the co-existence of different, even in mutually inconsistent models that are connected by semantic mappings. The nature of the proposed semantic mappings satisfies the requirements of the medical domain, because they do not require any changes to the connected ontologies and do not create logical inconsistency even if the models are incompatible.

This paper is organized as follows. We first briefly review the central definitions of the extended OWL semantics. In particular, we introduce the notion of local domains and mappings between them as well as their formal interpretation. In section 3 we describe the setting of a case study we conducted in using OWL to define and reason about alignments of medical ontologies and present some examples from the alignment. The use of C-OWL for reasoning about alignments is discussed in section 4. We conclude with a summary of our experiences and a discussion of the role of C-OWL for terminology integration in the medical domain.

2 Contextual semantics for OWL

The main idea of the proposed contextual semantics for OWL is split to up the global interpretation of different OWL ontologies into a set of local interpretations for each ontology. In order to make the alignment of ontologies with contradicting definitions possible, the notion of a hole is introduced which makes every statement in an ontology satisfiable. As a consequence statements are allowed to hold in one ontology but not in another one[1].

Definition 1 (OWL interpretation with local domains) *An* OWL interpretation with local domains *for a set of OWL ontologies* $\{\langle i, O_i\rangle\}_{i\in I}$*, is a family* $\mathcal{I} = \{\mathcal{I}_i\}_{i\in I}$*, where each* $\mathcal{I}_i = \left\langle \Delta^{\mathcal{I}_i}, (.)^{\mathcal{I}_i}\right\rangle$*, called the* local interpretation *of* O_i*, is either an interpretation of* O_i *into* $\Delta^{\mathcal{I}_i}$*, or a hole.*

The definition above completely separates the interpretations of different ontologies. As our aim is, however, to represent and reason about alignment between different ontologies, we have to introduce a way of connecting their domains. C-OWL does this by means of so-called bridge rules that define the semantic relations between concepts in different ontologies. C-OWL defines the following kinds of bridge rules stating that a concept from an ontology O_i is more general, more specific, equivalent, disjoint or overlapping with a concept from another ontology O_j:

$$i\!:\!x \xrightarrow{\sqsubseteq} j\!:\!y, \quad i\!:\!x \xrightarrow{\sqsupseteq} j\!:\!y, \quad i\!:\!x \xrightarrow{\equiv} j\!:\!y, \quad i\!:\!x \xrightarrow{\perp} j\!:\!y, \quad i\!:\!x \xrightarrow{*} j\!:\!y,$$

A mapping between two ontologies is a set of bridge rules between them. A context space is a pair composed of a set of OWL ontologies $\{\langle i, O_i\rangle\}_{i\in I}$ and a family $\{M_{ij}\}_{i,j\in I}$ of mappings from i to j, for each pair $i, j \in I$. To give the semantics of context mappings the definition of an OWL interpretation with local domains is extended with the notion of *domain relation*. A domain relation $r_{ij} \subseteq \Delta^{\mathcal{I}_i} \times \Delta^{\mathcal{I}_j}$ states, for each element in $\Delta^{\mathcal{I}_i}$ to which element in $\Delta^{\mathcal{I}_j}$ it corresponds to. The semantics for bridge rules from i to j can then be given with respect to r_{ij}. The interpretation for a context space is composed of an OWL interpretation with holes and local domains and the an interpretation domain relation from i to j, which is a subset of $\Delta^{\mathcal{I}_i} \times \Delta^{\mathcal{I}_j}$. As suggested above, the definition of bridge rules introduces semantic relationships between concepts in different ontologies thereby constraining

[1]For technical details about interpretations with holes see [Bouquet et al., 2003]

the global interpretation. As the way bridge rules are interpreted is important with respect to the possibilities for reasoning about alignments we give the formal definition of satisfiability of bridge rules.

Definition 2 (Satisfiability of bridge rules[2]) *Let $\mathfrak{I}$ be the global interpretation of a context space, then*

1. $\mathfrak{I} \models i{:}x \xrightarrow{\sqsubseteq} j{:}y$ *if* $r_{ij}(x^{\mathcal{I}_i}) \subseteq y^{\mathcal{I}_j}$;

2. $\mathfrak{I} \models i{:}x \xrightarrow{\sqsupseteq} j{:}y$ *if* $r_{ij}(x^{\mathcal{I}_i}) \supseteq y^{\mathcal{I}_j}$;

3. $\mathfrak{I} \models i{:}x \xrightarrow{\equiv} j{:}y$ *if* $r_{ij}(x^{\mathcal{I}_i}) = y^{\mathcal{I}_j}$;

4. $\mathfrak{I} \models i{:}x \xrightarrow{\perp} j{:}y$ *if* $r_{ij}(x^{\mathcal{I}_i}) \cap y^{\mathcal{I}_j} = \emptyset$;

5. $\mathfrak{I} \models i{:}x \xrightarrow{*} j{:}y$ $r_{ij}(x^{\mathcal{I}_i}) \cap y^{\mathcal{I}_j} \neq \emptyset$;

An interpretation for a context space is a model for it if all the bridge rules are satisfied.

3 Aligning Medical Ontologies: An Experiment in Using C-OWL

In the medical area a lot of work has been done on the definition and standardization of terminologies [3]. The result of these efforts is a large number of medical terminologies and classifications. The complexity of the terminologies used in medicine and the strong need for quality control has also lead to the development of ontologies that feature complex concept definition (compare [Golbreich et al., 2003] for a discussion of the required expressiveness). Some of these ontologies are available in OWL and can be seen as the first OWL applications that have a use in real life applications. C-OWL and especially its formal semantics provides us with several possibilities concerning the alignment of the medical ontologies mentioned above.

[3] see e.g. http://www.medinf.mu-luebeck.de/ ingenerf/terminology/Index.html for a collection of standards

3.1 Alignment Scenario

In our Case study, we used available representations of the the following medical ontologies:

Galen The Motivation for the GALEN project [Rector and Nowlan, 1993] is the difficulty in exchanging clinical data between different persons and organizations due to the heterogeneity of the terminology used. As a result of the project, the GALEN Coding Reference model has been developed. This reference model is an ontology that covers general medical terms, relations between those terms as well as complex concepts that are defined using basic terms and relations. We used an OWL version of the GALEN model that contains about 3100 classes and about 400 relations.

Tambis The aim of the Tambis [Baker et al., 1999] (Transparent Access to Bioinformatics Information Sources) is to provide an infrastructure that allows researchers in Bioinformatics to access multiple sources of biomedical resources in a single interface. In order to achieve this functionality, the project has developed the Tambis Ontology, which is an explicit representation of biomedical terminology. The complete version of Tambis contains about 1800 terms. The DAML+OIL version we used in the case study actually contains a subset of the complete ontology. It contains about 450 concepts and 120 Relations.

UMLS The Unified Medical Language System UMLS [Nelson and Powell, 2002] is an attempt to integrate different medical terminologies and to provide a unified terminology that can be used across multiple medical information sources. Examples of medical terminologies that ave been integrated in UMLS are MeSH and SNOWMED. In our case study, we used the UMLS semantic network. The corresponding model that is available as OWL file contains 134 semantic types organized in a hierarchy as well as 54 relations between them with associated domain and range restrictions.

We assume that the goal is to establish a connection between the Tambis and the GALEN ontology in such a way that the two models with their different focus supplement each other. An option for aligning Tambis and GALEN is an indirect alignment based on a third, more general model of the domain. In this setting the two models are made comparable by aligning each one with the third, more general model and using the semantic relations in this third model together with the mapping to determine the relation between classes in the two ontologies.

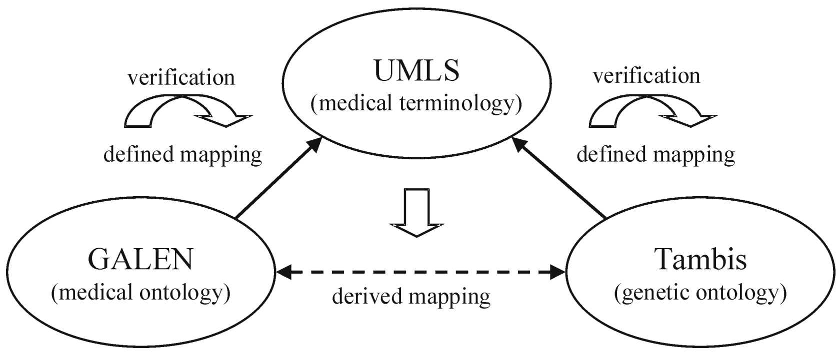

Figure 1: Indirect Alignment of Tambis and GALEN using UMLS

The UMLS semantic network is such a general model. Being the result of an integration of different medical terminologies(compare [Bodenreider, 2004]), we can assume that the network is general enough to cover the content of Tambis, GALEN and also other prospective ontologies that we might want to align. In order to explore the use of C-OWL for the alignment of medical ontologies, we conducted a small case study in aligning the ontologies mentioned above using the UMLS semantic network as a central terminology. We investigated the upper parts of the ontologies and identified areas with a sufficient overlap. Such an overlap between all three models exists with respect to the following three areas:

Processes: Different physiological, biological and chemical processes related to the functioning of the human body and to the treatment of malfunctions.

Substances: Substances involved in physiological processes including chemical, biological and physical substances.

Structures: Objects and object assemblies that form the human body or parts of it. Further, structures used in the treatment of diseases.

We analyzed the three models with respect to these three topics. Based on the comparison of the three models, we defined mappings between Tambis and GALEN and the UMLS terminology. These mappings consist of sets of bridge rules each connecting single concepts or concept expressions. In the following, we present some alignment examples from the case study. In particular we describe some of the alignment of GALEN and UMLS with respect to substances A more detailed description of the case study can be found in [Stuckenschmidt, 2004].

3.2 Examples from the Alignment

GALEN contains the notion of a generalized substance which is a notion of substance that subsumes substances in a physical sense and energy making it more general than the notion of substance in UMLS

$$\text{GeneralisedSubstance} \xleftrightarrow{\sqsupseteq} \text{Substance}$$

The actual notion of substance as defined in GALEN is not as we might expect equivalent to the notion of substance in UMLS, because it also contains some notions that are found under anatomical structures in UMLS. We can, however, state that the GALEN notion of substance is more specific than the union of substances and anatomical structures in UMLS.

$$\text{Substance} \xleftrightarrow{\sqsubseteq} \text{Substance} \sqcup \text{Anatomical\_Structure}$$

The next GALEN concept that also occurs in UMLS but has a slightly different meaning is the notion of body substance. The difference is illustrated in the fact that it also covers the notion of tissue which is found under anatomical structures in UMLS. We conclude that the notion of body substance in GALEN in a broader one than in UMLS.

$$\text{BodySubstance} \xleftrightarrow{\sqsupseteq} \text{Body\_Substance}$$

The other main class of substances mentioned in GALEN are chemical substances. Looking at the things contained under this notion, we conclude that it is equivalent to the notion of chemical in UMLS.

$$\text{ChemicalSubstance} \xleftrightarrow{\equiv} \text{Chemical}$$

We can also find the correspondences to the distinction between elementary and complex chemicals made by GALEN in UMLS. Elementary chemicals are a special case of the UMLS concept of elements ion or isotope.

$$\text{ElemetaryChemical} \stackrel{\sqsubseteq}{\longleftrightarrow} \text{Element\_Ion\_or\_Isotop}$$

Complex chemicals contain all kinds of chemical substances sometimes viewed structurally, sometimes functionally. Therefore, we cannot related this concept to one of these views taken by UMLS. We also notice that there are notions of complex chemicals in GALEN that do not occur under chemicals in UMLS - e.g. Drugs that related to the concept of clinical drug classified under manufactured objects.

$$\text{Drug} \stackrel{\equiv}{\longleftrightarrow} \text{Clinical\_Drug}$$

Further, the UMLS views on chemicals also contain elementary chemicals. Consequently, we can only define the notion of complex chemical to be compatible with the union of the two views in UMLS

$$\text{ComplexChemical} \stackrel{*}{\longleftrightarrow} \text{Chemical\_Viewed\_Structurally} \sqcup \text{Chemical\_Viewed\_Functional}$$

On the level of more concrete chemical notions we find a number of correspondences mentioned in the following. Named hormones are equivalent to hormones in UMLS

$$\text{NAMEDHormone} \stackrel{\equiv}{\longleftrightarrow} \text{Hormone}$$

Proteins are more specific than amino acids, peptides or proteins.

$$\text{Protein} \stackrel{\sqsubseteq}{\longleftrightarrow} \text{Amino\_Acid\_Peptide\_or\_Protein}$$

The notions of lipid and of carbohydrate are the same in the two models

$$\text{Lipid} \stackrel{\equiv}{\longleftrightarrow} \text{Lipid}$$

$$\text{Carbohydrate} \stackrel{\equiv}{\longleftrightarrow} \text{Carbohydrate}$$

There is an overlap between the notion of acid in GALEN and the concepts amino acid, peptide or protein and Nucleic acid , nucleosid or protein in UMLS.

$$\text{Acid} \xleftrightarrow{*} \text{Amino\_Acid\_Peptide\_or\_Protein} \sqcup \text{Nucleic\_Acid\_Nucleosid\_or\_Protein}$$

Finally metals can be defined to be a special case of inorganic chemicals.

$$\text{Metal} \xleftrightarrow{\sqsubseteq} \text{Inorganic\_Chemical}$$

In summary, we were able to find a lot of correspondences on the level of groups of chemicals. While the models disagreed on the higher level structuring of substances, they shared a lot of more concrete concepts. As a consequence, we found a number of equivalence and subsumption relationships between substances at a lower level while at the more general level, we often had to use weak relations or link to very general concepts.

4 Reasoning about Alignments

In the experiment, we defined mappings in a ad-hoc rather than a systematic fashion. Such an ad hoc approach for defining mappings bears the risk of inconsistency and in completeness. We cannot prevent the creation of inconsistent or incomplete mappings, but the semantics of C-OWL can be used to verify and extend a defined mapping in order to detect inconsistencies and implied mappings. In the following we give examples of the use of the C-OWL semantics to verify and extend the mappings between the substance information in the different medical ontologies.

4.1 Verification of Mappings

A mapping can become inconsistent if two classes who are known to overlap, e.g. because they are subclasses of each other, link to disjoint concepts in another model. An example of this situation can be found in the substance related part of the alignment between Tambis and UMLS. Figure 2 shows the situation. On the right hand side the extensions of the UMLS concept chemical substances and some of its subclasses are sketched. UMLS distinguishes between chemical from a structural and a functional view. In the case where these two views are defined to be disjoint (one can either take a structural or a functional view but not both) we get an inconsistency with the mappings defined for the Tambis ontology, because the mappings claims that the image of the concept chemical is exactly the extension of the structural view. At the same time, we claim that the image of enzyme which is a subclass of chemical is exactly the extension of the UMLS concept Enzyme which is classified under the functional view on chemicals in UMLS and therefore disjoint from the structural view. This however is now possible in the C-OWL semantics as the image of enzyme is a subset of the image of chemical by definition.

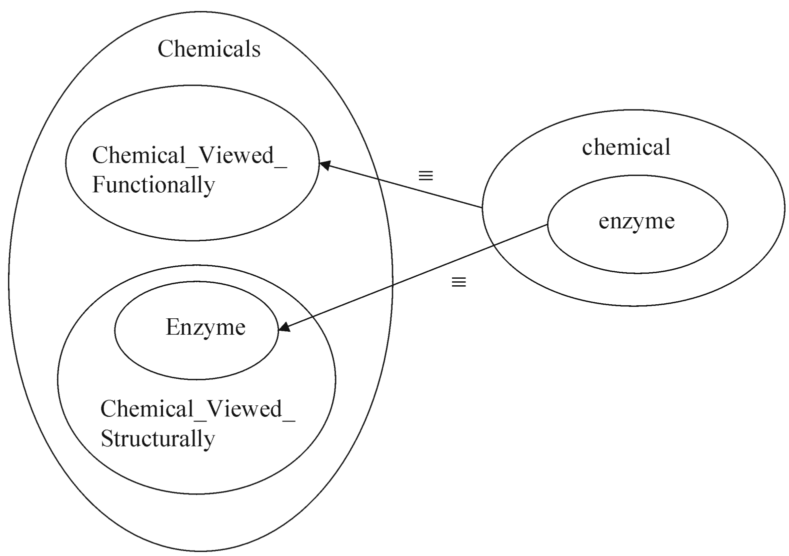

Figure 2: An Inconsistent Mapping

This ability to detect inconsistencies depends on the existence of appropriate disjointness statements in the ontology the mappings point to. Alternatively, the use of disjointness mappings can provide the same effect. If we want to make clear that chemicals in Tambis are not classified according to the functional view (which we just found to be not entirely true) we can also add a corresponding mapping stating that the image of chemicals is disjoint from the extension of the functional view on chemicals. The definition of this mapping will have the same effect leading to an inconsistency as described above.

4.2 Derivation of Semantic Relations

Besides the possibility to detect inconsistencies in the mappings, we can also infer additional bridge rules between the same models based on existing ones thereby making the complete mapping implied by the defined rules explicit. We illustrate this possibility by discussing possible implications of an equivalence mapping. Figure 3 illustrates parts of the alignment of substance related alignment of UMLS and GALEN. In particular, it shows the rule stating an equivalence between the GALEN class chemical and the UMLS class chemical substance which is part of the alignment. The definitions in UMLS state that chemical substances are less general than the class generalized substance, more general than complex chemicals and disjoint from processes. As the existing bridge rule states that the image of chemical is exactly the extension of chemical substance in UMLS, these relations also hold between this image and the other UMLS classes mentioned. The relations can be explicated by adding corresponding bridge rules stating that the image of chemicals is more general than complex chemicals, less general that generalized substance and disjoint from processes.

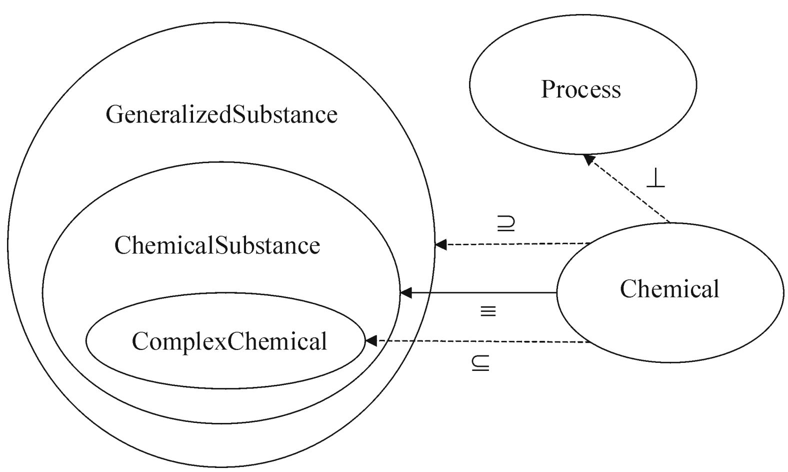

Figure 3: Derivation of additional Mappings

Similar inferences can be made based on bridge rules indicating specialization and generalization relations. If we replace the equivalence in figure 3 by a rule stating that chemicals is more specific than chemical substances, we are still able to infer the relations to generalized substances and to processes. Just the one to complex chemicals will be lost, because the image of chemicals might only overlap or be disjoint from the extension of the respective concept. Conversely, replacing the equivalence by bridge rule stating that chemicals is more general than chemical substances would have preserved the conclusion that chemicals is more general than complex chemicals. Finally, stating that chemicals is disjoint from chemical substances would have implied that it is also disjoint from complex chemicals.

4.3 Merging Local Models

Another thing we would like to do based on the alignments is to compare the local models (Tambis and GALEN) with each other and derive semantic correspondences between classes in these models as well. It turns out that we cannot really drive mappings between the two local models from their mappings to UMLS, because referring to different interpretation domains, we cannot compare the constraints imposed by these mappings. This situation changes, however, when we assume that the local models are to be merged. In this case, their interpretation domain becomes the same and we can use the constraints to derive semantic correspondences between concepts in the two models from the existing mappings.

Figure 4 shows two examples of derived relations between concepts from GALEN and Tambis. The figure shows two concepts from each, UMLS (upper part), Tambis (lower left part) and GALEN (lower right part). We assume that we have fixed the inconsistency detected in the mapping from Tambis to UMLS by removing the bridge rule relating chemical substances to the structural view on chemicals and replacing it by an equivalence between chemical substance and chemicals in general. As the GALEN concept chemical is also defined to be equivalent to Chemical, we can derive that these two concepts are equivalent in the merged ontology. Further, we defined the notion of substance in Tambis to be more specific than the same notion in UMLS which is again defined to be more specific than generalized substance in GALEN. From these mappings, we can derive that the Tambis notion of substance is more specific than Generalized substance and add a corresponding axiom to the merged ontology.

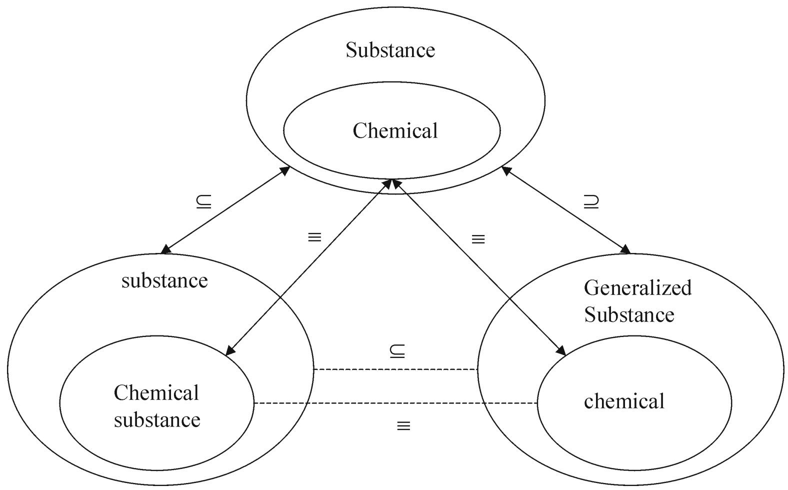

Figure 4: Derivation of semantic relations in the merged model

5 Discussion

We conclude that C-OWL provides a suitable formalism for supporting the alignment of complex terminologies like the ones we face in the medical area. While allowing the co-existence of different views, C-OWL still provides powerful reasoning support for the verification and derivation of mappings and even supports the process of merging terminologies based on existing mappings. These possibilities are essential for support knowledge engineers in the task of specifying mappings which currently mainly is a manual task. C-OWL is designed in such a way that no changes to existing OWL ontologies are required. Alignment mappings can be specified independently just referring to existing ontologies. This makes C-OWL directly applicable to existing ontologies like the ones mentioned in this paper. We are currently developing an RDF-based syntax for mapping definitions in C-OWL. The next steps of the developments of C-OWL is the develop of tools that support the creation, visualization and the reasoning about alignments.

References

[Baker et al., 1999] Baker, P., Goble, C., Bechhofer, S., Paton, N., Stevens, R., and Brass, A. (1999). An ontology for bioinformatics applications. *Bioinformatics*, 15(6):510–520.

[Bodenreider, 2004] Bodenreider, O. (2004). The unified medical language system (umls): integrating biomedical terminology. *Nucleic Acids Research*, 32.

[Bouquet et al., 2003] Bouquet, P., Giunchiglia, F., van Harmelen, F., Serafini, L., and Stuckenschmidt, H. (2003). C-OWL: Contextualizing ontologies. In Sekara, K. and Mylopoulis, J., editors, *Proceedings of the Second International Semantic Web Conference*, number 2870 in Lecture Notes in Computer Science, pages 164–179. Springer Verlag.

[Gangemi et al., 1998] Gangemi, A., Pisanelli, D., and Steve, G. (1998). Ontology integration: Experiences with medical terminologies. In Guarino, N., editor, *Proceedings of Conference on formal ontologies in information systems (FOIS'98)*. IOS Press.

[Golbreich et al., 2003] Golbreich, C., Dameron, O., Gibaud, B., and Burgun, A. (2003). Web ontology language requirements w.r.t expressiveness of taxonomy and axioms in medicine. In *The SemanticWeb - ISWC 2003*, volume 2870 of *Lecture Notes in Computer Science*, pages 180–194. Springer Verlag.

[Nelson and Powell, 2002] Nelson, S. and Powell, Tammy andn Humphreys, B. (2002). The unified medical language system (umls) project. In Kent, A. and Hall, C., editors, *Encyclopedia of Library and Information Science.*, pages 369–378. Marcel Dekker, Inc.

[Rector and Nowlan, 1993] Rector, A. and Nowlan, W. (1993). The galen project. *Computer Methods and Programs in Biomedicine*, 45:75–78.

[Stuckenschmidt, 2004] Stuckenschmidt, H. (2004). Using C-OWL for the alignment of medical ontologies a case study. Technical report, Vrije Universiteit Amsterdam.