International Journal on Semantic Web and Information Systems 2007;3(2):1-26.

Experience in Aligning Anatomical Ontologies

Songmao Zhang[1], Ph.D., Olivier Bodenreider[2], M.D., Ph.D.
[1]Institute of Mathematics, Academy of Mathematics and Systems Science
Chinese Academy of Sciences, Beijing, P. R. China
[2]U.S. National Library of Medicine, NIH, Bethesda, Maryland, USA
[1]smzhang@math.ac.cn, [2]olivier@nlm.nih.gov

ABSTRACT

An ontology is a formal representation of a domain modeling the entities in the domain and their relations. When a domain is represented by multiple ontologies, there is need for creating mappings among these ontologies in order to facilitate the integration of data annotated with these ontologies and reasoning across ontologies. The objective of this paper is to recapitulate our experience in aligning large anatomical ontologies and to reflect on some of the issues and challenges encountered along the way. The four anatomical ontologies under investigation are the Foundational Model of Anatomy, GALEN, the Adult Mouse Anatomical Dictionary and the NCI Thesaurus. Their underlying representation formalisms are all different. Our approach to aligning concepts (directly) is automatic, rule-based, and operates at the schema level, generating mostly point-to-point mappings. It uses a combination of domain-specific lexical techniques and structural and semantic techniques (to validate the mappings suggested lexically). It also takes advantage of domain-specific knowledge (lexical knowledge from external resources such as the Unified Medical Language System, as well as knowledge augmentation and inference techniques). In addition to point-to-point mapping of concepts, we present the alignment of relationships and the mapping of concepts group-to-group. We have also successfully tested an indirect alignment through a domain-specific reference ontology. We present an evaluation of our techniques, both against a gold standard established manually and against a generic schema matching system. The advantages and limitations of our approach are analyzed and discussed throughout the paper.

Keywords: Ontology, ontology alignment, knowledge representation, anatomy, Semantic Web

INTRODUCTION

An ontology is a formal representation of a domain modeling the things in that domain and the relationships between those things. Generally speaking, ontologies are composed of concepts (or classes) organized in taxonomies and other hierarchical structures, including partonomies. Moreover, concepts are often connected by various kinds of associative relationships (e.g., spatial, temporal, functional, etc.). In addition to relations to other concepts, concepts can be represented as having properties, often used to differentiate them from other concepts. An ontology, more formally, is "a set of logical axioms designed to account for the intended meaning of a vocabulary" (e.g., Guarino, 1998a). Different ontologies are created to support different tasks, including data integration (e.g., Goble et al., 2001), reasoning (e.g., Horrocks & Sattler, 2001) and the semantic annotation of resources in the Semantic Web (e.g., Kiryakov et al., 2003).

A given domain is often represented by multiple ontologies, providing overlapping, yet different coverage and possibly differing in their representation of the domain knowledge. There is a need for creating mappings among such ontologies in order

to facilitate the integration of data annotated with these ontologies and reasoning across ontologies. The goal of ontology alignment is to identify correspondences among entities (i.e., concepts and relationships) across ontologies with overlapping content. Manual alignment of large ontologies is slow, difficult, labor intensive and error prone. Moreover, it is not suitable for applications in which ontologies need to be aligned on the fly. Semi-automatic and fully automatic approaches to aligning ontologies have been developed instead.

Anatomy is central to the biomedical domain and many anatomical representations have been created over the past fifteen years. While some of them are mere lists of names for anatomical entities (e.g., Terminologica Anatomica), others are full-fledged ontologies, organizing anatomical entities in a rich network of relations. Different knowledge representation formalisms have been used to represent anatomical ontologies, including frame-based structures (e.g., the Foundational Model of Anatomy) and description logics (e.g., GALEN common reference model, SNOMED CT®). While most anatomical ontologies available represent human anatomy, the study of model organisms by biologists (Bard, 2005) has prompted the development of anatomical ontologies for other species (e.g., the Adult Mouse Anatomical Dictionary). Like domain ontologies in general, most anatomical ontologies are developed for a given purpose, for example to support cancer research (NCI Thesaurus) or clinical applications (SNOMED CT). In contrast, some ontologies, called reference ontologies, have been developed independently of specific objectives. For example, the Foundational Model of Anatomy, a reference ontology of structural anatomy (Rosse & Mejino, 2003), could be used as a reference for describing physiology and pathology.

Over the past few years, we have developed domain knowledge-based techniques for aligning large anatomical ontologies, with the objective of exploring approaches to aligning representations of anatomy differing in formalism, structure, and domain coverage. We started by aligning concepts point-to-point in two large ontologies of human anatomy, using lexical and structural techniques (Zhang & Bodenreider, 2003). We later tested these techniques on other pairs of anatomical ontologies, both within and across species (Bodenreider, Hayamizu, Ringwald, de Coronado, & Zhang, 2005; Bodenreider & Zhang, 2006). We also investigated the complex alignment of groups of concepts (Zhang & Bodenreider, 2006a) and that of relationships (Zhang & Bodenreider, 2004a). Finally, we investigated the possibility of deriving the indirect alignment of two ontologies through their direct alignment to a reference ontology (Zhang & Bodenreider, 2005). The objective of this paper is to recapitulate our experience in aligning anatomical ontologies and to reflect on some of the issues and challenges encountered along the way. In particular, we want to show the importance of domain-specific knowledge in our alignment strategies.

The paper is organized as follows. We first briefly review related work on ontology alignment. Then, we present our experience in aligning anatomical concepts directly, both point-to-point and group-to-group. We follow by the presentation of the alignment of relationships. Finally, we present the indirect alignment techniques we developed. The evaluation of our techniques is presented next, both against a gold standard established manually and against a generic schema matching system. The advantages and limitations of our approaches are analyzed and discussed throughout the paper.

BACKGROUND

The general framework of this study is that of ontology aligning, merging, matching, and integration. More than merging or integrating ontologies, i.e., transforming several source ontologies into a single ontology, we are interested in establishing a correspondence between equivalent entities across partially overlapping ontologies. This task is traditionally called mapping, matching or alignment. Ontology matching is an active field of research, supported by a dynamic community[1]. It is beyond the scope of this paper to give a detailed account of the various approaches proposed for aligning ontologies. We will rather outline salient aspects of

[1] See http://ontologymatching.org/ and http://www.atl.lmco.com/projects/ontology/ for an example of the resources created by this community.

such approaches as they relate to our work. For a detailed survey of such approaches, the interested reader is referred to several reviews published recently (Doan & Halevy, 2005; Kalfoglou & Schorlemmer, 2003; Noy, 2004a; Rahm & Bernstein, 2001; Shvaiko & Euzenat, 2005). Recent reviews of ontology matching tools include (Noy, 2004b) and papers contrasting existing tools to a particular one (e.g., Kotis, Vouros, & Stergiou, 2006). The rest of this section discusses some key features of alignment systems, including ours.

Rule-based vs. learning-based mapping. Our approach is entirely based on rules, some of which are specific to the domain. Compared to learning-based approaches, it would necessarily be more difficult to generalize to other domains.

Schema vs. ontology matching. While approaches to matching database schemas can generally be applied to ontology matching, a richer and more explicit semantics is usually found in ontologies. However, the semantics in most biomedical terminologies (or lightweight ontologies) is probably comparable to that of database schemas.

Schema vs. instance level. Ontology languages such as OWL can represent both classes and instances. In databases, the instances correspond to the data content, i.e., values found in the columns in the database. In anatomical ontologies, however, the anatomical entities represented correspond essentially to classes, not instances[2]. Therefore, the methods we use operate essentially at the schema level.

Granularity of mappings. Because the ontologies we map represent similar domains, albeit across species in some cases, our goal is to establish correspondences between concepts at the same level of granularity. In other words, we are mostly interested in identifying equivalent concepts across ontologies. As a consequence, when ontologies of different granularities are compared, the finer-grained concepts in one ontology might not be mapped to concepts in the other, even though suitable subsumers would exist in the other ontology.

Mapping cardinality. Most of the mappings we identify are point-to-point (1:1) mappings between concepts. However, we also investigated complex rules for aligning concepts group-to-group (1:n and n:m). Finally, we also found it useful to identify those concepts provably without mappings (1:0), including fine-grained concepts in one ontology with no equivalent in a coarser ontology. When mapping relationships, we report both 1:1 and 1:n mappings.

Lexical techniques. Like most systems, we use the lexical properties of concept names for the mapping. However, we differ from these systems in many respects. The names of anatomical entities present in ontologies are essentially noun phrases, simple (e.g., *First tarsometatarsal joint*) or including prepositional clauses (e.g., *Neck of femur*). Instead of using syntactic information, lemmatization, partial matches and edit distance to model lexical resemblance, we rely on a linguistically-motivated model of term variation specifically developed for the biomedical domain (McCray, Srinivasan, & Browne, 1994). In practice, we seek exact and normalized matches between terms. Normalization makes the input and target terms potentially compatible by eliminating such inessential differences as inflection, case, underscore and hyphen variations, as well as word-order variation. Two terms are considered lexically equivalent when they have the same normalized form and lexically different otherwise. Whenever available in the ontologies, synonyms are used in addition to preferred terms in order to determine concept similarity at the lexical level.

Structural techniques. Like most systems also, we use the structural properties of the ontology for the mapping, namely the existence of shared relations across ontologies. Hierarchical relations (taxonomic and partitive) constitute the backbone of anatomical ontologies and have therefore a better chance of being represented consistently across ontologies. For this reason, we apply structural techniques only on hierarchical relations. We use structural techniques essentially for validating the matches obtained at the lexical level. In other words, for two concepts c_1 and c_1' to be equivalent across ontologies, they first need to have equivalent names, but also to share relations to other

[2] In ontology parlance, instances correspond not to leaf nodes, but to actual entities in reality (e.g., *my* liver), as opposed to the class *liver*.

equivalent concepts, e.g., to c_2 and c_2', respectively. By sharing relations, what we mean is that there needs to be paths between c_1 and c_2 and between c_1' and c_2', respectively. However, the paths are not required to be identical across ontologies. What distinguishes our approach from others is that we use domain knowledge to make explicit the relations that would otherwise not be represented in the ontologies. The complementation, augmentation and inference techniques we use are presented later with the methods. Making relations explicit can be understood as normalizing relations across ontologies in order to facilitate the structural comparison, of which it represents a critical element. We also use the structural features of ontologies to derive group-to-group mappings.

External resources. Several systems use WordNet[3], the electronic lexical database for the English language, as a source of lexical knowledge (e.g., synonyms) and domain knowledge. The biomedical equivalent of WordNet is the Unified Medical Language System®[4] (UMLS®), whose Metathesaurus® comprise 1.3 million concepts and some 5 million names (Bodenreider, 2004). Metathesaurus concepts are the equivalent of WordNet synsets in the sense that synonymous terms are clustered together to form the list of names for a concept. Rather than single-word terms as it is mostly the case in WordNet, the Metathesaurus comprises mostly complex, multi-word terms, suitable for mapping the complex terms found in biomedical ontologies.

Semantic constraints. The use of semantic constraints is relatively limited in our approach[5], especially because there is less need for term disambiguation in the narrow domain of anatomy than in a more general context. In fact, we use semantic constraints (e.g., disjointness among top-level classes) when mapping between ontologies whose content is not specific to anatomy (e.g., GALEN). In this case, the semantic incompatibility between classes is used to prevent lexically similar (ambiguous) terms from being mapped.

[3] http://wordnet.princeton.edu/
[4] http://umlsks.nlm.nih.gov/
[5] Arguably, the structural requirements themselves already constrain the semantics.

Automatic vs. interactive mapping. Our approach is automatic and requires no input from the user. For this reason, the structural techniques used to validate the mappings suggested lexically are conservative, calibrated to yield a minimal number of false positives. In our experience, about 10 percent of the mappings are not supported by structural evidence and would therefore not be suggested by the automatic system, but could be reviewed by domain experts for accuracy.

Reference ontologies. The role of reference ontologies in ontology alignment is mentioned in several systems (e.g., Kalfoglou & Schorlemmer, 2002). However, in these systems, the goal is to generate an isomorphism between local ontologies (populated with instances by different communities) and a reference ontology (unpopulated). In contrast, we propose to map the "local ontologies" not only to the reference, but also to themselves, *through* the reference. More formally, we use direct mappings of two ontologies O_1 and O_2 to a reference domain ontology O_r to derive an indirect mapping between O_1 and O_2. More recently, (Aleksovski, Klein, ten Kate, & van Harmelen, 2006) and (Aleksovski, ten Kate, & van Harmelen, 2006) also used background knowledge to match two biomedical onlologies with limited overlap.

The many ontology alignment systems available include PROMPT (Noy & Musen, 2000), CUPID (Madhavan, Bernstein, & Rahm, 2001), FCA-Merge (Stumme & Maedche, 2001), HCONE-Merge (Kotis, Vouros, & Stergiou, 2006), and GLUE (Doan, Madhavan, Domingos, & Halevy, 2004). With AnchorPrompt (Noy, 2004b), we share the notion of "anchor" (i.e., a pair of related terms across ontologies, established by lexical similarity in our case) and the use of shared paths between anchors across ontologies to validate the similarity among related terms. Therefore, AnchorPrompt is undoubtedly the system to which our approach is the most closely related. The major differences between AnchorPrompt and our approach can be summarized as follows. AnchorPrompt creates a sophisticated similarity score based on path length and other features. In contrast, we use a simpler validation scheme based on paths restricted to combinations of taxonomic and

partitive relations, suitable for the anatomical domain. Unlike AnchorPrompt, our approach does not rely on path length and is therefore less sensitive to differences in granularity between ontologies. Both Anchor-Prompt and our approach identify equivalence relations between groups of concepts. In Anchor-Prompt, concepts in such groups must be linked by taxonomic relations, whereas this is not a requirement in our approach. Other features of our approach not found in Anchor-Prompt include mapping non-anchors to groups of anchors and identifying concepts provably without mapping in the other ontology.

In summary, our approach to aligning concepts (directly) is automatic, rule-based, and operates at the schema level, generating mostly point-to-point mappings. It uses a combination of domain-specific lexical techniques (to map entities at the element, not instance level) and structural and semantic techniques (to validate the mappings suggested lexically). It also takes advantage of domain-specific knowledge (lexical knowledge from external resources such as the UMLS, as well as knowledge augmentation and inference techniques). Additionally, we have successfully tested an indirect alignment through a domain-specific reference ontology. The contribution of this paper, rather than producing new alignment techniques or tools, is to adapt existing techniques to the specific domain of anatomy and to apply and evaluate these techniques to the mapping of large-scale anatomical ontologies, both within and across species.

MATERIALS

We give a brief overview of the four ontologies used in our mapping experiments. Two of them (the Foundational Model of Anatomy and the Adult Mouse Anatomical Dictionary) are pure anatomical ontologies, while the other two (GALEN and the NCI Thesaurus) are broader biomedical ontologies of which anatomy represents a subdomain. Although more recent versions of some of these resources are available, we refer to the older versions presented below throughout this paper in order to facilitate comparisons across our own studies.

The **Foundational Model of Anatomy**[6] (FMA) is an evolving ontology that has been under development at the University of Washington since 1994 (Noy, Musen, Mejino, & Rosse, 2004; Rosse & Mejino, 2003). Its objective is to conceptualize the physical objects and spaces that constitute the human body. The underlying data model for the-FMA is a frame-based structure implemented with Protégé[7]. 71,202 concepts cover the entire range of macroscopic, microscopic and subcellular canonical anatomy. In addition to preferred terms (one per concept), 52,713 synonyms are provided (up to 6 per concept). For example, there is a concept named *Uterine tube*, which has two synonyms: *Oviduct* and *Fallopian tube*. Because single inheritance is one of the modeling principles used in the FMA, every concept (except for the root) stands in a unique *IS-A* relation to other concepts. Additionally, seven kinds of partitive relationships are used to connect anatomical concepts (e.g., *part of*, *constitutional part of*, *regional part of*, and their inverses *part*, *constitutional part*, *regional part*). Beside hierarchical relationships, there are 81 kinds of associative relationships between concepts in the FMA. While most of them have inverses (e.g., *branch of* and *branch*), a few do not (e.g., *input from*). The version used in this study was downloaded on December 2, 2004.

The Generalized Architecture for Languages, Encyclopedias and Nomenclatures in medicine[8] (**GALEN**) has been developed as a European Union AIM project led by the University of Manchester since 1991 (Rector et al., 1997; Rogers & Rector, 2000). The GALEN common reference model is a clinical terminology based on description logics. GALEN contains 25,322 concepts and intends to represent the biomedical domain, of which canonical anatomy is only one part. Only one name is provided for each non-anonymous concept (e.g., *Lobe of thyroid gland*). There are 3,170 anonymous concepts (e.g., `SolidStructure which <isPairedOrUnpaired leftRightPaired>`). GALEN supports multiple inheritance and every concept in GALEN (except for the root) stands in

6 http://fma.biostr.washington.edu/
7 http://protege.stanford.edu/
8 http://www.opengalen.org/

at least one – and often several – *IS-A* relations to other concepts. Relationships in GALEN are generally finer-grained than in the FMA. There are 41 kinds of *PART-OF* relationships (e.g., *isStructuralComponentOf, IsDivisionOf)*, and 536 associative relationships (e.g., *isBranchOf*, *isServedBy*). All relationships have inverses (e.g., *hasStructuralComponent*, *HasDivision*, *hasBranch*, *serves*). The version used in this study is version 6 of the Common Reference Model.

The **Adult Mouse Anatomical Dictionary** (MA)[9] is a structured controlled vocabulary describing the anatomical structure of the adult mouse (Hayamizu, Mangan, Corradi, Kadin, & Ringwald, 2005). It comprises 2,404 concepts. Each concept has one name (e.g., *Head muscle* and *Adrenal artery*). Additionally, 240 concepts have a total of 259 synonyms (e.g., *Limb* has synonym *Extremity*). The ontology is represented as a directed acyclic graph whose edges represent the relationships *IS-A* and *PART-OF*. Every concept is connected to other concepts through *IS-A* or *PART-OF* relationships. However, about 38% of the concepts do not have any *IS-A* relationship to other concepts (e.g., *Knee PART-OF Hindlimb* is the only hierarchical relation available for *Knee)*. On the other hand, nearly 4% of the concepts have more than one *IS-A* relationship to other concepts (e.g., *Hand phalanx* is both a kind of *Phalanx* and *Hand digit bone*). The version used in this study was downloaded on December 22, 2004 (under the name Mus adult gross anatomy in the Open Biomedical Ontologies[10]).

The **NCI Thesaurus** (NCI) [11] provides standard vocabularies for cancer research (De Coronado, Haber, Sioutos, Tuttle, & Wright, 2004) and its anatomy class describes naturally occurring human biological structures, fluids and substances. The ontology is available in the Ontology Web Language (OWL DL). There are 4,410 anatomical concepts (accounting for about 12% of all NCI concepts). Every concept has one preferred name (e.g., *Abdominal esophagus*). Additionally, 1,207 concepts have a total of 2,371 synonyms (e.g., *Orbit* has synonym *Eye socket*). Except for the root (*Anatomic Structure, System, or Substance*), every anatomical concept has at least one *IS-A* relationship to another concept, and nearly 4% of the concepts have more than one *IS-A* relationship to other concepts (e.g., *Radius bone* is both a kind of *Long bone* and *Bone of the upper extremity*). In addition, anatomical concepts are also connected by a *PART-OF* relationship (named *Anatomic Structure Is Physical Part of*). The version used in this study is version 04.09a (September 10, 2004).

DIRECT ALIGNMENT

In our approach to aligning two ontologies directly, we first identify similar concepts point-to-point across ontologies using lexical and structural techniques. Then, based on this point-to-point alignment, additional complex mappings among groups of concepts are identified solely on the basis of structural features. Finally, the associative relationships across ontologies are also compared, again based solely on structural information.

The alignment between the FMA and GALEN is used as the main case study in this section, but we also present the results of the alignment of another pair of anatomical ontologies: the Adult Mouse Anatomical Dictionary and the NCI Thesaurus.

For alignment purpose in this study, we considered as only one *PART-OF* relationship (with *HAS-PART* as its inverse) the various kinds of partitive relationships present in the FMA and GALEN.

Aligning concepts point-to-point

We identified one-to-one concept mappings between the FMA and GALEN using lexical resemblance between concept names and then validated the mappings through shared hierarchical paths among concepts across ontologies.

Lexical alignment

The lexical alignment identifies shared concepts across ontologies based on lexical similarity between concept names. For the FMA, both preferred concept names and synonyms are used in the lexical alignment process. For GALEN, only non-

[9] http://www.informatics.jax.org/searches/anat-dict_form.shtml
[10] http://obo.sourceforge.net/
[11] http://cancer.gov/cancerinfo/terminologyresources/

anonymous concept names are used. Lexical similarity is assessed through exact match and after normalization. The normalization program distributed with the UMLS provides a linguistically-motivated model for lexical resemblance adapted to the specificity of biomedical terms, abstracting away from minor differences in terms including case, hyphen, inflection and word order variations (McCray, Srinivasan, & Browne, 1994).

Concepts exhibiting similarity at the lexical level across ontologies are called anchors, as they are going to be used as reference concepts in the structural validation and for comparing associative relationship. Additional anchors are identified through synonymy in an external resource: the Unified Medical Language System (UMLS). More specifically, two concepts across ontologies are considered anchors if their names are synonymous in the UMLS Metathesaurus (i.e., if they name the same concept) and if the corresponding concept is in the anatomy domain (i.e., has a semantic type related to *Anatomy*).

Examples of anchors, shown in Figure 1, include the concepts *Cardiac valve* in the FMA and *Valve in heart* in GALEN, identified as anchor concepts because *Cardiac valve* has *Valve of heart* as a synonym in the FMA and *Valve in heart* matches *Valve of heart* after normalization. Additionally, *Fibrous ring of mitral valve* (with synonym *Mitral anulus*) in the FMA and *Mitral ring* in GALEN form an anchor because *Mitral anulus* and *Mitral ring* are synonyms, i.e., they are both names for the concept *Structure of anulus fibrosus of mitral orifice* in the UMLS.

3,431 matching anchor concepts were identified lexically, accounting for about 4.8% of the FMA concepts and 13.5% of GALEN concepts. 328 out of 3,431 anchors were identified through UMLS synonymy.

Structural validation

In the structural validation of the lexical alignment, the first step is to acquire the semantic relations explicitly represented in the ontologies. Inter-concept relationships are generally represented by semantic relations <c_1, r, c_2>, where the relationship r links concepts c_1 and c_2. Because they form the backbone of anatomical ontologies and are therefore more likely to be represented consistently across ontologies, hierarchical relationships only are considered at this step. These relationships are *IS-A* and *PART-OF*, along with their inverses *INVERSE-IS-A* and *HAS-PART*, respectively. Having extracted the relations explicitly represented in the ontologies, we then normalize the representation of the relations in each ontology in order to facilitate structural comparisons across ontologies. We first complement the hierarchical relations represented explicitly with their inverses as necessary. Implicit semantic relations are then extracted from concept names (augmentation) and various combinations of hierarchical relations (inference). Augmentation and inference are the two main techniques used to acquire implicit knowledge from the FMA and GALEN. For a detailed analysis of the contribution of each technique, the interested reader is referred to (Zhang & Bodenreider, 2004b).

Complementation. As partial ordering relationships, hierarchical relationships are anti-symmetric. However, *IS-A* and *PART-OF* have inverse relationships, *INVERSE-IS-A* and *HAS-PART*. Except for *IS-A*, not every relation is represented bidirectionally. For example, <*External ear, HAS-PART, External acoustic tube*> is explicitly represented in the FMA but its inverse relation is missing. In canonical anatomy, the inverse relations are essentially always valid, although this is not necessarily the case in the real world (Smith et al., 2005). For the sole purpose of aligning ontologies, in order to facilitate the comparison of paths between anchors across ontologies, we complement the FMA and GALEN with the inverse relations that are not explicitly represented. For example, we generated the relation <*External acoustic tube, PART-OF, External ear*>.

Augmentation attempts to represent with relations knowledge that is otherwise embedded in the concept names. Augmentation is based on linguistic phenomena, such as the reification of partitive relations. In this case, a relation <***P***, *PART-OF*, ***W***> is created between concepts ***P*** (the part) and ***W*** (the whole) from a relation <***P***, *IS-A*, ***Part of W***>, where the concept ***Part of W*** reifies, i.e., embeds in its name, the *PART-OF* relationships to ***W***. For example, <*Neck of femur, PART-OF, Joint*> was added from the relation <*Neck of femur, IS-A, Component of joint*>,

where the concept *Component of joint* reifies a specialized *PART-OF* relationship. Examples of augmentation based on other linguistic phenomena include <*Sweat gland, IS-A, Gland*> (from the concept name *Sweat gland*) and <*Extensor muscle of leg, PART-OF, Leg*> (from the concept name *Extensor muscle of leg*). The semantics of nominal modification generally corresponds to subsumption (e.g., the head noun *gland* modified by *sweat* is a hypernym of *gland*). In contrast, the semantics of prepositional clauses introduced by *of* is not necessarily a partitive relation (e.g., *glass of wine* is not part of *wine*). Here, domain knowledge was required to assess what relations can be automatically extracted with high accuracy in the particular context of anatomical terms. We determined that partitive relations could be accurately created from prepositional clauses introduced by *of* in anatomical terms containing no other prepositions.

Inference generates additional semantic relations by applying inference rules to the existing relations in order to facilitate the comparison of paths between anchors across ontologies. These inference rules, specific to this alignment, represent limited reasoning along the *PART-OF* hierarchy, generating a partitive relation between a specialized part and the whole or between a part and a more generic whole. For example, <*First tarsometatarsal joint, PART-OF, Foot*> was inferred from the relations <*First tarsometatarsal joint, IS-A, Joint of foot*> and <*Joint of foot, PART-OF, Foot*>. Analogously, <*Interphalangeal joint of thumb, PART-OF, Finger*> was inferred from the relations <*Interphalangeal joint of thumb, PART-OF, Thumb*> and < *Thumb, IS-A, Finger*>. The number of hierarchical and partitive relations extracted and generated is listed in Table 1. Not surprisingly, many relations come from inference, which performs similarly to a transitive closure of the hierarchical relations.

Table 1. Number of relations in the FMA and GALEN

Types of relations	FMA	GALEN
Explicitly represented	238,641	123,069
Complemented	167,381	18,955
Augmented	162,392	25,916
Inferred	5,559,762	1,235,070
Total	6,128,176	1,403,010

With these explicit and implicit semantic relations, the structural validation identifies structural similarity and conflicts among anchors across ontologies. Structural similarity, used as **positive structural evidence**, is defined by the presence of common hierarchical paths among anchors across ontologies, e.g., <c_1, *PART-OF*, c_2> in one ontology and <c_1', *PART-OF*, c_2'> in another where $\{c_1, c_1'\}$ and $\{c_2, c_2'\}$ are anchors across ontologies[12]. The anchor concepts *Cardiac valve* in the FMA and *Valve in heart* in GALEN, presented earlier, received positive structural evidence because they share hierarchical paths to some of the other anchors across ontologies. For example, as illustrated in Figure 1, *Cardiac valve* is related to *Heart* (*PART-OF*), to *Mitral valve* (*INVERSE-IS-A*) and to *Mitral ring* (*HAS-PART*).

Conflicts, on the other hand, are used as **negative structural evidence**. The first type of conflict is defined by the existence of hierarchical paths between the same anchors across ontologies going in opposite directions, e.g., <c_1, *PART-OF*, c_2> in one ontology and <c_1', *HAS-PART*, c_2'> in the other. The second type of conflict is based on the disjointness of top-level categories across ontologies (i.e., semantic constraints). For example, *Nail* in the FMA is a kind of *Skin appendage* which is an *Anatomical structure*, while *Nail* in GALEN is a *Surgical fixation device* which is an *Inert solid structure*. *Anatomical structure* and *Inert solid structure* being disjoint top-level categories, the two concepts *Nail* in the FMA and GALEN are semantically distinct, which prevents them from being aligned although they have exactly the same name.

Table 2 shows the results of structural validation where anchors are classified into three sets with respect to the kind of structural evidence exhibited.

Anchors with no structural evidence. 9.9% of anchors do not receive any structural evidence. For example, although linked to *Myocyte* (*HAS-PART*) and *Muscle* (*IS-A*) in GALEN, *Supinator muscle* has no connections to other anchors in the FMA. The absence of *any* paths to other anchors represents about two thirds of the cases. The remaining cases

[12] The transitive closure of hierarchical relation greatly facilitates paths comparison across ontologies, because complex paths between anchors are represented by a single relation.

correspond to the absence of *shared* paths to other anchors across ontologies. For example, although *Venule* is linked to thirteen anchors in the FMA (e.g., *Basement membrane*, *Lipid*), and five in GALEN (e.g., *Blood vessel, Cardiovascular system*), none of these paths are shared across ontologies.

Anchors with positive structural evidence. 88.8% of all anchors receive positive evidence, most of them sharing hierarchical paths of the same type (e.g., *Cardiac valve* in the FMA and *Valve in heart* in GALEN, presented earlier). An example of shared "compatible" hierarchical relations is the anchor *Pelvic fascia*. In both ontologies, this concept is linked to *Visceral pelvic fascia*, but, although going in the same direction, the relationship is *INVERSE-ISA* in GALEN and *HAS-PART* in the FMA. For alignment purposes, sharing compatible hierarchical relations is deemed a sufficient condition.

Anchors with negative structural evidence. 1.3% of the anchors represent conflicts between the two ontologies. For example, the relationship between the anchors *Apex of bladder* and *Urinary bladder* is *PART-OF* in GALEN but *HAS-PART* in the FMA. Another type of conflict is represented by the semantic incompatibility between *Nail* (the anatomical structure) in the FMA and *Nail* (the medical device used to treat fractures) in GALEN presented earlier.

Overall, starting from the 3,431 possible anchors and excluding 44 pairs exhibiting negative evidence as well as 188 cases of ambiguous mapping (disambiguated manually), the lexical alignment followed by structural validation finally identified 3,199 pairs of equivalent concepts in the FMA and GALEN, accounting for about 4% of all FMA concepts and 13% of all GALEN concepts. (The limited overlap between the two ontologies is discussed is the section titled *Concepts provably without matches* below).

The same alignment technique was applied to another pair of anatomical ontologies: the Adult Mouse Anatomical Dictionary (MA) and the NCI Thesaurus (NCI). Table 3 shows the result of the relation acquisition process. Of note, in these two ontologies, hierarchical relations are always represented unidirectionally. This is why the number of relations complemented corresponds exactly to the number of relations represented explicitly. Another difference with the FMA-GALEN alignment is that we did not extract additional relations from MA and NCI terms (augmentation), because the terms were generally less complex and included few embedded relations. Overall, as shown in Table 4, the lexical alignment followed by structural validation identified 715 pairs of equivalent concepts between the Adult Mouse Anatomical Dictionary (MA) and the NCI Thesaurus (NCI), accounting for about 30% of all MA concepts and 30% of those 2400 NCI concepts representing anatomical entities at a similar level of granularity. The proportion of lexical matches supported by positive structural evidence is roughly the same (about 90%) in the FMA-GALEN and MA-NCI alignments. Of note, no negative structural evidence was identified for any of the anchors in the MA-NCI alignment.

Table 2. Results of structural validation for the FMA-GALEN alignment

Structural evidence		3,431 anchors		
No evidence	No paths to other anchors	190	340	9.9%
	No shared paths to other anchors	150		
Positive evidence	Shared paths to other anchors(same type)	2065	3047	88.8%
	Shared paths to other anchors ("compatible")	982		
Negative evidence	Conflicting paths to other anchors	26	44	1.3%
	Semantic disjointness	18		

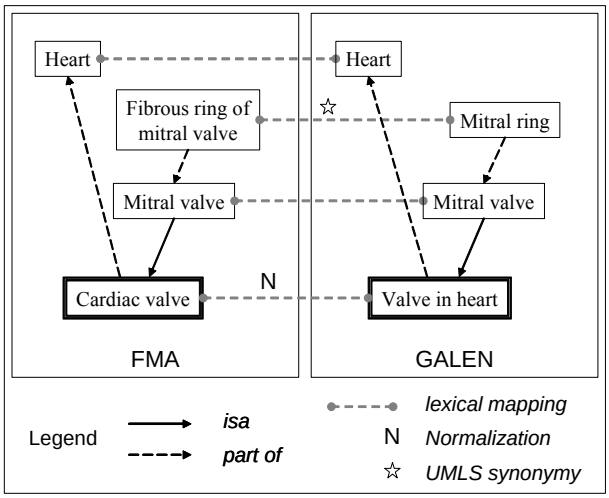

Figure 1. Structural validation following lexical alignment

Table 3. Number of relations in MA and NCI

Types of relations	MA	NCI
Explicitly represented	2,926	7,250
Complemented	2,926	7,250
Augmented	0	0
Inferred	15,044	45,302
Total	20,896	59,820

Table 4. Results of structural validation for the MA-NCI alignment

Structural evidence		715 anchors		
No evidence	No paths to other anchors	44	62	8.7%
	No shared paths to other anchors	18		
Positive evidence	Shared paths to other anchors(same type)	580	653	91.3%
	Shared paths to other anchors ("compatible")	73		
Negative evidence	Conflicting paths to other anchors	0	0	0%
	Semantic disjointness	0		

Aligning concepts group-to-group

Using the lexical alignment method followed by structural validation presented above, 3,199 pairs of equivalent concepts were identified between the FMA and GALEN, accounting for about 4% of the FMA concepts and 13% of GALEN concept. The complex structural rules introduced here allowed us to identify additional mappings and to identify concepts for which it can be demonstrated that no mapping to the other ontology can be found. Overall, about 44% of the FMA concepts and 69% of GALEN concepts were characterized in the alignment. In what follows, the term *anchor* refers to the 3,199 one-to-one matches obtained previously. Those are represented by double-lined boxes in figures. In contrast, the other concepts in the two ontologies are *non-anchors* (represented by single-lined boxes in figures). Finally, $anchor_D(X)$ denotes the set of all anchors in the descendants of concept X.

One-to-group matches

The following rules were developed for identifying matches in two different circumstances: 1) between non-anchors concepts, and 2) between non-anchor concepts in one ontology and anchors in the other.

Mapping between non-anchor concepts. Let us consider the two non-anchors X_1 and X_2 in one ontology and the non-anchor Y in another ontology. If $anchor_D(X_1)$ and $anchor_D(X_2)$ are not subsets of each other, and $anchor_D(X_1) \cup anchor_D(X_2) = anchor_D(Y)$ holds, then it is possible that a single concept Y matches a group of concepts $\{X_1, X_2\}$. For example as shown in Figure 2, the non-anchor concept *Extremity long part* in GALEN has four anchors in its descendants: *Arm*, *Forearm*, *Leg* and *Thigh*. In the FMA, the non-anchor concept *Proximal free limb segment* has two anchors in its descendants: *Arm* and *Thigh*, and *Middle free limb segment* has two other anchors in its descendants: *Forearm* and *Leg*. The set of anchors among the descendants of *Extremity long part* is thus the union of the sets of anchors in the descendants of *Proximal free limb segment* and *Middle free limb segment*. Therefore, we suggest a one-to-group match between the concept *Extremity long part* in GALEN and the group of concepts {*Proximal free limb segment*, *Middle free limb segment*} in the FMA. 22 such one-to-group matches were identified, corresponding to 36 non-anchors in the FMA and 30 in GALEN.

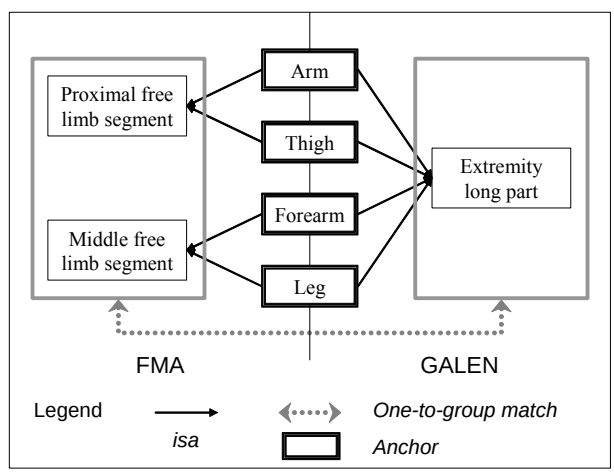

Figure 2. One-to-group match between Extremity long part in GALEN and {Proximal free limb segment, Middle free limb segment} in the FMA

Mapping between non-anchor concepts in one ontology and anchors in the other. In addition to the mappings between non-anchors presented above, one-to-group mappings can also occur be-

tween one non-anchor in one ontology and a group of anchors in the other ontology. This is often due to the use of different modeling principles in the two ontologies.

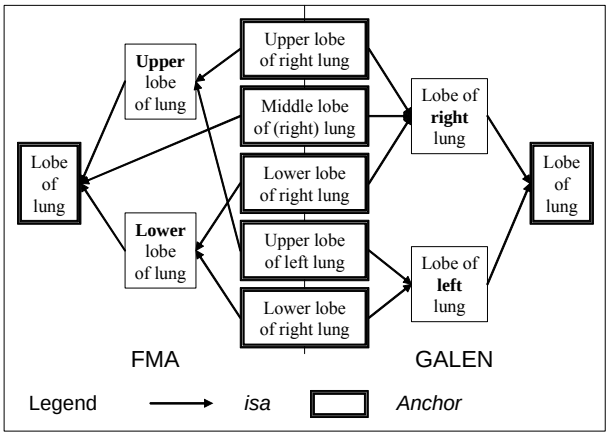

Figure 3.Differences in the representation of lobes of lung between the FMA and GALEN.

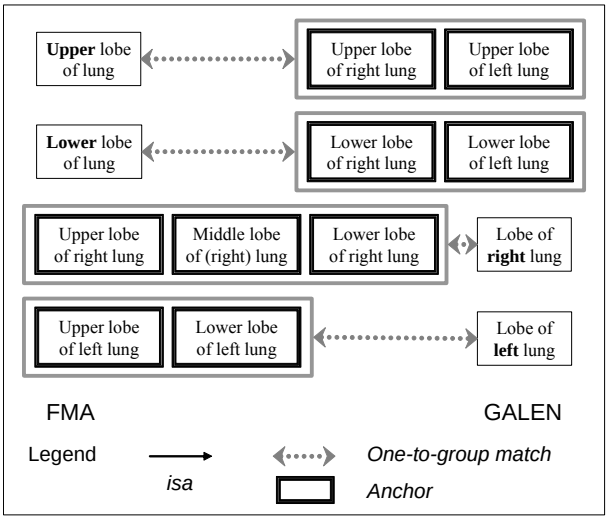

Figure 4. One-to-group matches in the descendants of Lobe of lung in the FMA and GALEN

For example, as illustrated in Figure 3, *Lobe of lung* [13] in the FMA is first modeled by upper/middle/lower position (i.e., *Upper lobe of lung*, *Middle lobe of lung* and *Lower lobe of lung*) and then by laterality (e.g., for *Upper lobe of lung*: *Upper lobe of left lung* and *Upper lobe of right lung*). By contrast, in GALEN, *Lobe of lung* is first modeled by laterality and then by upper/middle/lower position. Our point-to-point alignment identified five anchors in the descendants of *Lobe of lung*. In addition, we identified four one-to-group matches across ontologies, shown in Figure 4. 49 such mappings between a non-anchor and a group of anchors were found, where 25 are one GALEN non-anchor matching FMA anchors, and 24 one FMA non-anchor matching GALEN anchors.

Group-to-group matches

Let us consider the concepts ***X*** and ***Y*** across ontologies forming an anchor. If ***X*** and ***Y*** share exactly the same set of anchors (possibly empty) in their children, and ***X*** and ***Y*** have the same number of non-anchors in their children: $\{X_1, \ldots, X_n\}$ and $\{Y_1, \ldots, Y_n\}$, respectively, then there is a possible mapping between the two groups of non-anchors, i.e., between $\{X_1, \ldots, X_n\}$ and $\{Y_1, \ldots, Y_n\}$. For example, the anchor *Anterior intercostal artery* in the FMA has eleven children and all of them are non-anchors: *First anterior intercostal artery* to *Eleventh anterior intercostal artery*. In contrast, the eleven non-anchor children of *Anterior intercostal artery* in GALEN are anonymous: *(AnteriorIntercostalArtery which <isSpecificallyNonPartitivelyContainedIn First IntercostalSpace>)* to *(AnteriorIntercostalArtery which <isSpecifically NonPartitivelyContainedIn EleventhIntercostalSpace>)*. These two groups of eleven non-anchors were mapped across ontologies. 49 such group-to-group matches were identified between the FMA and GALEN, involving 127 non-anchors in each ontology.

Some inaccurate group-to-group mappings were identified, often related to differences in modeling between ontologies. For example, the anchor *Head of radius* in GALEN has two non-anchor children: *Distal head of radius* and *Proximal head of radius*, while the anchor *Head of radius* in the FMA has two non-anchor children: *Head of left radius* and *Head of right radius*. A group-to-group match was identified between the groups *{Distal head of radius, Proximal head of radius}* in GALEN and *{Head of left radius, Head of right radius}* in the FMA. However, this group-to-

[13] The right lung comprises three lobes: upper, middle and lower. The left lung has only two lobes: upper and lower.

group mapping is invalid because the two groups of children are based on two different classificatory principles. Distal/Proximal refers to the position in reference to the center of the body, while left/right refers to laterality. Such inaccurate mappings suggest that group-to-group mappings should be reviewed systematically by a domain expert.

Concepts provably without matches

The total number of concepts in the FMA is about three times of that in GALEN. Intuitively, there should be a large number of FMA concepts either mapping to GALEN concepts group-to-one, or simply having no matches in GALEN (e.g., because of differences in granularity between the FMA and GALEN). For example, the anchor *Submucosa* is a leaf node in GALEN, while it has 128 descendants in the FMA. All of its descendants are non-anchors and represent specialized concepts specific to the FMA, e.g., the submucosa of various organs. These 128 non-anchors were identified as having no matches in GALEN. Overall, 1,482 such cases were found, involving 11,189 FMA non-anchors and accounting for about 16% of all the FMA concepts.

On the other hand, some high-level concepts in GALEN represent non-canonical anatomical categories (e.g., *Non normal phenomenon*), anatomy-related categories (e.g., *Process, Graft*), and non-anatomical categories (e.g., *Food, Risk factor*). The concepts subsumed by these categories in GALEN are not expected to have matches in the FMA which is solely concerned with canonical (i.e., "normal") anatomical entities. 13,626 such non-anchor concepts in GALEN were identified (2,051 of them are anonymous), accounting for 53.8% of all GALEN concepts. Examples include *Supernumerary thumb* as a descendant of *Non normal phenomenon*, and the anonymous concept `(Alcohol which <playsPhysiologicalRole FoodRole>)` under *Food*.

Other structural matches

The structural techniques we developed for identifying matches to groups of concepts can also be applied to identifying point-to-point matches among non-anchor concepts. Two non-anchors ***X*** and ***Y*** across ontologies are likely to be a match if they reach the same non-empty set of anchors in their descendants, i.e., $\boldsymbol{anchor_D(X) = anchor_D(Y)}$. For example, as shown in Figure 5, the non-anchor *Cuneiform* in GALEN (with three leaf anchor descendants) and the non-anchor *Cuneiform bone* in the FMA (with three non-leaf anchor descendants) were identified as a match, because they both share the three anchors found in their descendants: *Medial cuneiform bone*, *Lateral cuneiform bone* and *Intermediate cuneiform bone*. 124 such one-to-one matches were found across ontologies.

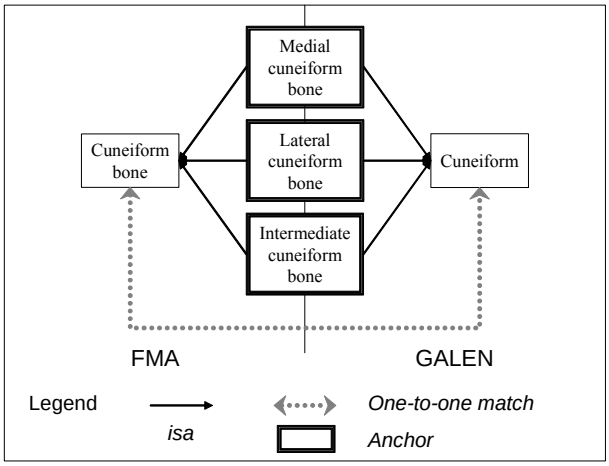

Figure 5. One-to-one match based on similar anchors in the descendants of Cuneiform bone in the FMA and Cuneiform in GALEN

Aligning relationships

While hierarchical relationships have been the object of careful inventory and standardization (Guarino, 1998b; Smith et al., 2005; Winston, Chaffin, & Herrmann, 1987), associative relationships tend to differ from ontology to ontology in names, semantics, and the constraints associated with their use. This is also because, unlike taxonomy and mereology which are required in virtually all ontologies, the theories expressed through associative relationships are generally specific to a subdomain. But even in a given subdomain, large differences may be observed in the use of associative relationships across ontologies, often corresponding to modeling choices.

We assume that one associative relationship in one ontology can be expressed in another ontology by either another associative relationship or a combination of associative and hierarchical relationships. We further assume the frequency with which a correspondence between two associative relationships is found to be a surrogate for the validity of the correspondence.

As anchors represent the correspondence between equivalent concepts across ontologies, relationship patterns represent the correspondence between relationships (or combination thereof) across ontologies. Such patterns are identified by investigating the relationships among anchors in the two ontologies. More precisely, for each associative relationship between two anchors in one ontology, we searched for all shortest paths between the same two anchors in the other ontology. Both hierarchical and associative relationships are allowed in the paths. However, we ignored the paths where an associative relationship and its inverse are present because such paths are usually not indicative of an associative relation of interest between the two anchors. For example, *Liver* → *isServedBy* → *Autonomic nerve of abdomen* → *serves* → *Surface of liver* was ignored for this reason. An associative relationship between two anchors in one ontology and a combination of relationships between the same two anchors in the other ontology compose a path pair.

Concepts are removed from the paths to create relationship patterns. Additionally, these patterns are simplified by representing several successive relationships of the same kind by only one relationship. These transformations generate pattern pairs from path pairs. For example, from the path pair:

FMA: *Pancreas* → *arterial supply* → *Dorsal pancreatic artery*

GALEN: *Pancreas* → *isServedBy* → *Caudal pancreatic artery* → *isBranchOf* → *Inferior pancreatic artery* → *isBranchOf* → *Dorsal pancreatic artery*

The following pattern pair is obtained:

FMA: *arterial supply*

GALEN: *isServedBy - isBranchOf*

This pattern pair is indirect as it involves more than one relationship. It would be direct otherwise. The frequency of each pattern pair (i.e., the number of paths pairs this pattern pair comes from) was recorded in order to select only the most frequent pairs (as they are also expected to be the most significant ones), thus ignoring "accidental" pattern pairs.

7,116 inter-anchor path pairs between the FMA and GALEN were obtained. 767 pattern pairs were identified from these path pairs, and 58 of them are direct pattern pairs. Table 5 lists some examples of pattern pairs.

Table 5. Example of pattern pairs

FMA	GALEN	Frequency
branch of	*isBranchOf*	364 (5 %)
PART-OF	*isBranchOf*	306 (4 %)
tributary of	*isBranchOf*	106 (1.5%)
HAS-PART	*isTo*	104 (1.5%)
member of	*ISA*	44 (0.6%)
arterial supply of - contained in	*isNonPartitively ContainedIn*	26 (0.4%)
contained in	*PART-OF - definesSpace*	2 (0.03%)

A small number of patterns occur with a high frequency, while the majority of patterns occur much less frequently (often only once or twice). The pattern pair with the highest frequency is *{FMA: branch of, GALEN: isBranchOf}*, and 364 path pairs have this pattern resulting in a frequency of 5% of the total 7,116 path pairs. The pattern pair with the second highest frequency (4%) is *{FMA: PART-OF, GALEN: isBranchOf}*, shared by 306 path pairs.

Table 6 illustrates the three types of patterns identified: one associative relationship corresponds to another associative relationship in another ontology (in 3% of the cases), indicative of equivalent associative relationships; one associative relationship corresponds to a combination of hierarchical and associative relationships in another ontology (in 92% of the cases), indicative of different levels of granularity or modeling choices in the two ontologies; and pattern pairs consist of one associative relationship in one ontology and a hierarchical relationship in the other (in 5% of the cases), where the two relationships in the pair must not be interpreted as being semantically equivalent.

Table 6. Analysis of relationship patterns

Types of patterns	Number of patterns (N = 767)		Examples
Associative corresponds to Associative	20	3%	F: *tributary of* G: *ISBRANCHOF*
Associative corresponds to Combination	709	92%	F: *arterial supply* G: *isServedBy – IS-A*
Associative and Hierarchical	38	5%	F: *bounded by* G: *HAS-PART*

Although we did not use lexical methods to match associative relationships, it is interesting to compare the results of our method to that of lexical techniques. Four occurrences are of similar relationship names and semantics, e.g., *{FMA: branch of, GALEN: isBranchOf}*. This represents the most straightforward case. Ten occurrences are of similar relationship names and differing semantics, e.g., no paths were found to support the pattern *{FMA: bounded by, GALEN: isSpaceBoundedBy}* that could have been suggested lexically. This illustrates why we did not want to rely on lexical similarity for matching relationships. Sixteen occurrences are of different relationship names and similar semantics, e.g., *{FMA: nerve supply, GALEN: isServedBy}*. This mapping would have been missed by methods relying solely on lexical similarity.

Lastly, although 81 associative relationships are defined in the FMA, only 47 of them were actually used to connect concepts in the ontology. A majority (43) of these relationships were identified in the alignment. However, this is not the case in GALEN. 534 out of 536 associative relationships were actually used, while no match was found for 431 of them (81%), e.g., *isPositionedInferiorTo*, *isExposedTo*, and *hasInternalExternalSelector*.

INDIRECT ALIGNMENT (THROUGH A REFERENCE)

Mappings among ontologies can be built pairwise, i.e., an alignment is created between every two ontologies. Alternatively, one ontology can be selected as the reference for mapping. All other ontologies only need to be mapped to this reference ontology and the pairwise mappings can be derived from the mappings to the reference ontology. These two approaches to aligning multiple ontologies are illustrated in Figure 6.

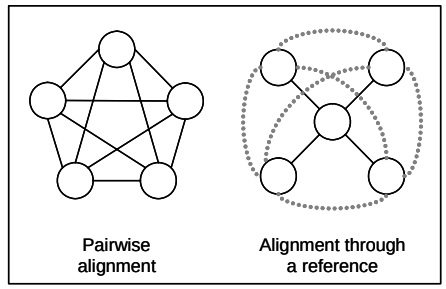

Figure 6. Aligning multiple ontologies

In this section, we compare two approaches to aligning multiple ontologies: pairwise ontology alignment and alignment through a reference ontology. We compare the direct alignment between two ontologies O_1 and O_2 to the indirect alignment automatically generated from mapping both O_1 and O_2 to O_R, the reference ontology. In practice, we perform: 1) three direct alignments O_1-O_2; O_1-O_R and O_2-O_R; 2) the indirect alignment between O_1 and O_2 through their direct alignments with O_R; and 3) a comparison of the direct alignment O_1-O_2 to the indirect alignment obtained through O_R. The creation of pairwise mappings among ontologies draws on the lexical alignment followed by structural validation described earlier. The three ontologies under investigation are the Foundational Model of Anatomy (FMA), the Adult Mouse Anatomical Dictionary (MA), and the anatomy subset of the NCI Thesaurus (NCI). Each of them was selected as a reference in turn to derive indirect matches between the other two.

Figure 7 shows an example of indirect alignment, where the FMA serves as a reference. The direct alignment MA-FMA identifies the match {MA: *Forelimb*, FMA: *Upper limb* (synonym: *Forelimb*)}, which is supported by positive evidence. The direct alignment NCI-FMA identifies the match {NCI: *Upper extremity*, FMA: *Upper limb* (synonym: *Upper extremity*)}, also supported by positive evidence. Therefore, the match {MA: *Forelimb*, NCI: *Upper extremity*} is derived automatically, through the FMA concept *Upper limb*, supported by positive structural evidence in both direct alignments.

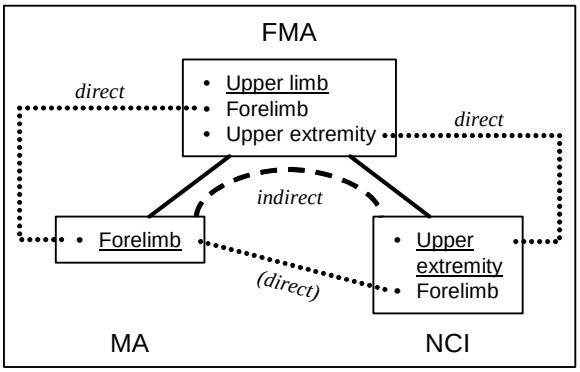

Figure 7. Indirect MA-NCI alignment through the FMA

Results for three direct alignments are summarized in the second row in Table 7 (DIR). The alignment NCI-FMA yielded the largest number of matches (2,173) and MA-NCI the smallest (715). A very small number of conflicts was identified in the two direct alignments to the FMA; none in the direct MA-NCI alignment. In the three direct alignments, a vast majority of the matches (> 90%) was supported by positive structural evidence. No evidence (positive or negative) was found for 5-9% of the matches in three direct alignments.

Table 7. Three direct vs. indirect alignments

	MA - NCI	**MA - FMA**	**NCI - FMA**
Direct alignment (DIR)	715 matches (91.3% pos. evidence)	1,353 matches (94.8% pos. evidence)	2,173 matches (90.1% pos. evidence)
Indirect alignment (IND)	FMA as reference	NCI as reference	MA as reference
	703 matches (92% pos. evidence)	771 matches (88.1% pos. evidence)	741matches (87.6% pos. evidence)
Shared by DIR & IND.	654 matches	708 matches	710 matches
Specific to DIR	61 matches	645 matches	1,463 matches
Specific to IND.	49 matches	63 matches	31 matches
Shared / DIR.	91.5%	52.3%	32.7%

Results for three indirect alignments are summarized in Table 7 (IND). For example, 703 matches between MA and NCI were automatically derived from direct alignments MA-FMA and NCI-FMA. 649 of them (92%) received positive structural evidence in both direct alignments MA-FMA and NCI-FMA, 8 (1%) received negative evidence in one of the two direct alignments, and 46 (7%) received no evidence in at least one of the two direct alignments.

Taking the three ontologies pairwise, we compared the matches obtained in their direct alignment to the matches resulting from their indirect alignment through the reference. The results of these comparisons are summarized in Table 7 (Shared by DIR & IND). For example, for the MA-NCI mapping (first column), 654 matches are shared by both alignments, leaving 61 matches specific to the direct alignment (accounting for 8.5% of the direct matches) and 49 specific to the indirect alignment through the FMA. Among the 654 shared matches, 583 (89%) received positive structural evidence in all three direct alignments (e.g., {MA: *Forelimb*, NCI: *Upper extremity*}).

About 10% of the shared matches in the three groups received no evidence in at least one of the three direct alignments. For example, although linked to other matches in MA (e.g., *HAS-PART Lung*) and the FMA (e.g., *HAS-PART Ear*), *Body* has no hierarchical relations to any other matches in NCI. This is why the matches of *Body* receive no evidence in the two direct alignments MA-NCI and NCI-FMA, while receiving positive evidence in direct alignment MA-FMA. On the other hand, nearly 1% of the shared matches in the three groups received negative evidence in one of the three direct alignments. For example, although a concept *Nephron* exists in the three ontologies, the corresponding match received negative evidence in the direct MA-FMA alignment (i.e., links to *Renal tubule* (synonym: *Uriniferous tubule*) through *HAS-PART* in MA but links to *Uriniferous tubule* through *PART-OF* in the FMA), while receiving positive evidence in both direct alignments MA-NCI and NCI-FMA. Domain knowledge is required to evaluate the match in these cases.

This study confirms the feasibility and efficiency of the indirect alignment through a reference ontology. Using the FMA as a reference resulted in the identification of a vast majority (91.5%) of the direct matches between MA and NCI. Moreover, the indirect alignment was able to identify matches not discovered by direct align-

ment. The large size of the FMA and its comprehensive set of synonyms contributed to this high percentage of mappings.

In contrast, when using NCI or MA as the reference in indirect alignment, only one half (52.3%) and one-third (32.7%), respectively, of the corresponding direct matches were identified. These findings confirm our intuition that ontologies offering a small number of concepts and a limited number of names for each concept are less suitable as a reference for deriving an indirect alignment between two ontologies.

Nevertheless, regardless of its size, as shown in Table 7 (Specific to IND), every ontology contributes specific indirect matches, i.e., matches that are not identified in the direct alignment. For example, using MA as a reference generated 31 specific matches, of which 19 received positive evidence in both direct alignments.

In summary, this study confirms that both the number of concepts and the number of concept names in the reference ontology are important parameters determining the suitability of an ontology to serve as a reference for deriving indirect mappings. These findings are compatible with Burgun's desiderata for domain reference ontologies in biomedicine, including good lexical coverage, good coverage in terms of relations and compatibility with standards (Burgun, 2005).

EVALUATION

The evaluation of mappings is still an open issue, especially for large-scale, domain specific ontologies represented in different formalisms, such as those discussed in this paper. One approach to evaluating ontology alignment is through competitive evaluation. Competitions such as the KDDCup[14] have been organized for almost a decade for data mining and knowledge discovery tasks. A similar effort started in 2004 for ontology alignment with the Ontology Alignment Evaluation Initiative[15] (OAEI), whose goal is to "establish a consensus for evaluation of ontology alignment methods", through the organization of annual challenges. Interestingly, beside general resources (e.g., web directories), anatomical ontologies – namely the FMA and GALEN – have been the object of the OAEI challenge in 2005 and 2006 (Euzenat, Stuckenschmidt, & Yatskevich, 2005). In 2005, two teams participated in the anatomy challenge (Jian, Hu, Cheng, & Qu, 2005; Kalfoglou & Hu, 2005). Their reports essentially outline the difficulties encountered along the way, including the large size of the anatomical ontologies and the transformation of both ontologies from their native format into OWL Full. Four teams contributed anatomy mappings in 2006.

One issue for the organizers of such competitive evaluations – and, more generally, for us in the evaluation of our methods – is the absence of a gold standard or ground truth, i.e., a list of mappings against which to compute, for example, precision and recall. Establishing such a gold standard for large, specialized ontologies is labor intensive and costly as it requires domain specialists (here anatomists) with some understanding of how anatomical knowledge is represented in each ontology (e.g., the use of metaclasses in Protégé for the FMA or anonymous concepts in GALEN).

Because we did not have enough resources for creating a gold standard for the FMA-GALEN alignment, we partnered with another team who had aligned the same ontologies using a generic schema matching system. For the smaller MA-NCI alignment, a biologist established a mapping manually between the two ontologies, to which we compared our direct alignment. Finally, beyond the proof of concept, the indirect MA-NCI alignment through the FMA used as a reference ontology can also be used for the purpose of evaluating the direct alignment.

Against another system

In parallel to our effort to align the FMA and GALEN at the National Library of Medicine, but unrelated to it, another alignment of the same ontologies was performed by Mork et al. at Microsoft Research (Mork, Pottinger, & Bernstein, 2004). Both teams had great difficulties for evaluating the alignments they produced. In 2004, we partnered

[14] http://www.acm.org/sigs/sigkdd/kddcup/index.php
[15] http://oaei.ontologymatching.org/

with the team at Microsoft and set out to align the same versions of the two ontologies[16] and to compare our results (Zhang, Mork, & Bodenreider, 2004). Although not ideal, this cross-validation provided some insights about the strengths and limitations of each approach.

In contrast to our knowledge-rich approach to aligning ontologies, Mork adapted a generic schema matching algorithm (Cupid) to cope with the large number of concepts in the ontologies and to handle the more expressive modeling environments. The matching algorithm operated in three successive phases: lexical, structural and hierarchical. For a detailed description of their approach, the interested reader is referred to (Mork & Bernstein, 2004). In contrast to ours, this system produces continuous similarity values between concepts. A similarity score higher than or equal to the threshold of .83 (determined heuristically) was required for matches identified lexically to be supported structurally.

The concept matches obtained in each alignment are summarized in Table 8. 2,776 concept matches were identified by Zhang et al. and 3,654 by Mork et al. Among them, a majority (2,199) both received positive structural evidence and had a similarity score above the threshold of .83, as shown in the upper left part of Table 8. These matches are supported by both alignments. Among the concept matches identified by Zhang and supported by positive structural evidence, 42 received similarity scores lower than the threshold and 295 were not identified by Mork. Conversely, among the concept matches identified by Mork and having a similarity score above the threshold of .83, 168 were supported by no structural evidence, 36 denoted conflicts (negative evidence), and 132 were not identified by Zhang. A detailed analysis of the differences between the two alignments is provided in (Zhang, Mork, & Bodenreider, 2004).

Although this simple cross-validation does not provide an absolute evaluation of recall (both approaches may fail to identify some mappings) or precision (both approaches could wrongly identify a match), it was reassuring to assess that most mappings were identified in common. Analyzing the discrepancies between the two alignments provided some insights about the limitations of each approach. For example, Mork identified 36 matches that denoted conflicts in Zhang, because their schema matching algorithm does not take advantage of semantic constraints. Conversely, because of lack of relations being represented in the FMA or GALEN, Zhang did not identify 168 matches for which no structural evidence could be found, most of which were valid matches identified by Mork.

Table 8. Concept matches in the two alignments (FMA-GALEN)

			Mork et al.		
			Identified		Not identified
			Similarity ≥ .83	Similarity < .83	
Zhang et al.	Identified	Positive evidence	2,199	42	295
		No evidence	168	*3*	*29*
		Negative evidence	36	*0*	*4*
	Not identified		132	*1,074*	

Against a gold standard established manually

As part of the caBIG project[17], the Jackson Laboratory created a mapping between the Adult Mouse Anatomical Dictionary (MA) and anatomical concepts in the NCI Thesaurus (NCI). This mapping between human and mouse anatomies is a critical resource for comparative science as diseases in mice are used as models of human disease. This mapping was created manually by a domain expert familiar with both human and mouse anatomy. All matches were validated using available anatomy resources. The mapping established by the expert constitutes a gold standard against

[16] The versions of the anatomical ontologies aligned by both teams in this comparison are slightly older than those presented in the Materials (FMA: dated of July 2, 2002; GALEN: Core Reference Model v. 4, dated of April 10, 2001).

[17] https://cabig.nci.nih.gov/

which we can evaluate our automatic alignment approach.

However, gold standards established by experts are not always completely accurate. Instead of using this gold standard mechanically to compute the traditional precision and recall values, we elected to use it for cross-validation purposes. In practice, our evaluation procedure was the following. We applied the lexical techniques presented earlier to the two ontologies and obtained a set of MA-NCI anchors. Then, we applied the structural techniques to find structural evidence in support of these mappings. We applied the same structural techniques to the set of gold standard anchors established manually by the expert in order to check their validity. Finally, we compared the manual mapping (with structural validation) to the automatic lexical mapping (with structural validation).

The concept matches obtained in each alignment are summarized in Table 9. 715 concept matches were identified by the automatic alignment (Zhang et al.) and 781 by the manual alignment used as our gold standard. Among them, a majority (594) received positive structural evidence in both alignments. These matches are the matches in common. Among the concept matches identified by Zhang and supported by positive structural evidence, 59 were not identified in the gold standard. Conversely, among the concept matches identified in the gold standard and supported by positive structural evidence, 132 were not identified by Zhang. A detailed analysis of the differences between the two alignments is provided in (Bodenreider, Hayamizu, Ringwald, de Coronado, & Zhang, 2005).

Table 9. Automatic alignment (Zhang et al.) vs. gold standard alignment (MA-NCI)

			Gold standard			
			Identified			Not ident.
			Pos. ev.	No ev.	Neg. ev	
Zhang et al.	Identified	Positive evidence	594	0	0	59
		No evidence	2	43	0	17
		Negative evidence	0	0	0	0
	Not identified		132	10	0	

In the series of experiments reported in this paper, this evaluation is the only one conducted against a gold standard established manually by a domain expert. Again, it was reassuring to assess that we identified over 75% of the mappings in the gold standard. Analyzing the 132 mappings missed by the automatic alignment revealed some limitations of our lexical techniques. Conversely, because it takes advantage of the rich set of synonyms provided by the UMLS, the automatic alignment identified 59 additional matches, deemed valid by the expert, but missed while establishing the gold standard.

Against an indirect alignment through a reference ontology

Although not completely independent, the direct and indirect alignments of a pair of ontologies can also provide some elements of cross-validation. As an illustration, we will use the alignments (presented earlier) performed between the Adult Mouse Anatomical Dictionary (MA) and anatomical concepts in the NCI Thesaurus (NCI) using the direct approach and the indirect mapping through the Foundational Model of Anatomy (FMA) used as a reference ontology.

The concept matches obtained in each alignment are summarized in Table 10. 715 concept matches were identified by the automatic alignment (Zhang et al.) and 703 by the indirect alignment through the FMA. Among them, a majority (583) received positive structural evidence in both alignments. Those are the matches in common. Among the concept matches identified in the direct alignment and supported by positive structural evidence, 53 were not identified in indirect alignment. Conversely, among the concept matches identified in indirect alignment and supported by positive structural evidence, 45 were not identified in the direct alignment. A detailed analysis of the differences between the two alignments is provided in (Zhang & Bodenreider, 2005).

The analysis of the differences between the direct and indirect alignments revealed that the presence of additional synonyms and relations in the FMA was responsible for the mappings identified

specifically be the indirect alignment. Conversely, differences in coverage and knowledge representation between the FMA on the one hand and MA and NCI on the other were responsible for the mappings identified specifically be the direct alignment. However, the most important finding of this study was that deriving an indirect alignment through a reference ontology was not only feasible, but also reasonably efficient.

Table 10. Direct vs. indirect alignment (MA-NCI)

			Indirect			
			Identified			Not ident.
			Pos. ev.	No ev.	Neg. ev	
Direct	Identified	Positive evidence	583	14	3	53
		No evidence	21	30	3	8
		Negative evidence	0	0	0	0
	Not identified		45	2	2	

CONCLUSIONS

This paper discusses the approach we developed for aligning concepts and relationships between large-scale anatomical ontologies. While the only mapping relation studied is equivalence, various types of mappings are considered: point-to-point, one-to-group and group-to-group. Additionally, we also identify the concepts provably without mapping to the other ontology. Our approach was tested on several pairs of anatomical ontologies, represented in various native formalisms, and was evaluated both against a gold standard alignment established manually and against a generic schema-matching approach.

One key feature in our approach to aligning domain ontologies is the use of domain knowledge throughout the alignment process. The lexical alignment relies on a model of lexical resemblance specifically developed for biomedical terms in the UMLS. The structural similarity takes advantage of implicit relations embedded in concept names made explicit by our augmentation techniques, as well as disjointness axioms added to the ontologies. Finally, domain knowledge is required for the evaluation of the alignment, i.e., to analyze the discrepancies between the system under investigation and the reference. We showed that domain knowledge was the main factor behind the identification of additional mappings by our approach compared to the generic schema matching approach.

Another important decision we made early on was to require no specific knowledge representation formalism for the ontologies to be aligned. For example, the FMA was represented in the frame-based system Protégé, while the representation of GALEN was based on the description logic language GRAIL. Instead of reducing both ontologies to a given formalism, we reduced them to their lowest common denominator: the simple triple-based representation illustrated earlier. Transforming the FMA to GRAIL, GALEN to frames or both of them to OWL, for example, would have been a nontrivial endeavor. Unrelated to the alignment, we later converted the FMA into OWL DL and concluded that this conversion required not only syntactic transformation, but also semantic enrichment (Zhang, Bodenreider, & Golbreich, 2006). In contrast, the organizers of the OAEI challenge elected to convert the FMA and GALEN to OWL Full. In our opinion, the resulting representation is both confusing and not suitable for supporting mapping experiments. The apparent "mismatch" between *Pancreas* in the FMA and GALEN once transformed into OWL Full, reported in (Kalfoglou & Hu, 2005), is an illustration of the distortion introduced by the transformation. The representation of *Pancreas* in the native environments is clearly indicative of a match.

As mentioned earlier, in order to facilitate the structural validation of the lexical alignment, we developed various techniques for making explicit the relations embedded in concept names and inferable from combinations of relations. For example, while applying our augmentation technique to the FMA (to generate <***P***, *PART-OF*, ***W***> from <***P***, *IS-A*, ***Subdivision of W***>), *Surface of umbilicus* was found in the descendants of *Subdivision of surface of umbilicus*, which is incorrect. We reported inconsistent representations to the developers of the FMA, who corrected the errors. We also formalized our consistency checking procedures into quality assurance guidelines for ontologies (Zhang & Bodenreider, 2006b).

Many issues are common to both ontology alignment and terminology alignment, including lexical mismatches, differences in scope and granularity, and versioning issues (Klein, 2001). Terminology integration in the UMLS relies heavily on lexical knowledge. Structural constraints are not enforced, by design, because the objective is simply to represent, not curate, terminological assertions. Limited semantic constraints are provided by the Semantic Network. The alignment process is only partially automated: while lexical similarity is used to determine candidate synonyms among over five million terms, term and concept properties are reviewed manually by the Metathesaurus editors (Bodenreider, 2004). In contrast, in our ontology alignment approach, structural similarity between ontology is essential for validating the lexical similarity between concept names. Our approach to aligning concepts point-to-point is mostly automatic in the sense that only those mappings not supported by positive or negative structural evidence need to be reviewed for accuracy. They usually represent about 10% of the mappings.

Despite the progress made in the past years in developing ontology alignment methods and tools, many unresolved issues remain and call for further research. In the future, we would like to exploit associative relations for mapping purposes, in addition to hierarchical relations. The limited success of alignment techniques based solely on structural features shows that names are usually represented more consistently than relations. Exploiting an ontology of relationships would certainly help make better use of structural features in ontologies for alignment and other purposes. The large proportion of concepts still uncharacterized after the alignment suggests that using only equivalence relations between ontologies might be too restrictive. Depending on the context of use, identifying subsumption relations might be of interest (e.g., for indexing or annotation purposes). More research is needed to determine the degree to which idioms in representation formalisms and differences in modeling impair the mapping to other ontologies. For example, SNOMED CT uses a representation of anatomical entities based on Structure-Entire-Part (SEP) distinctions (Schulz & Hahn, 2005). In the SEP framework, the right hand (*Entire right hand*) is represented as follows:

- *<Entire right hand isa Entire hand>*
- *<Entire right hand isa Structure of right hand>*
- *<Entire right hand part_of Entire right upper extremity>*

Although not entirely intuitive, this representation offers interesting computational properties derived from the reification of part of relations. On the other hand, such a representation has significant consequences on the alignment of SNOMED CT with other anatomical ontologies (Bodenreider & Zhang, 2006). Finally, the validity of any mapping must be examined in light of the purpose for which it has been developed. In practice, the level of precision and granularity required in a given application determines in part the validity of the mapping. For example, while the mapping of *Prostate* between human and mouse anatomies seems plausible, given that both male humans and mice have prostates, it does not account, for example, for the fact that the mouse has five prostates, where the human has one (Travillian, Gennari, & Shapiro, 2005), because this information is simply not represented in most anatomical ontologies.

ACKNOWLEDGEMENTS

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and by the Natural Science Foundation of China (No.60496324), the National Key Research and Development Program of China (Grant No. 2002CB312004), the Knowledge Innovation Program of the Chinese Academy of Sciences, MADIS of the Chinese Academy of Sciences, and Key Laboratory of Multimedia and Intelligent Software at Beijing University of Technology.

Thanks for their support and encouragement to Cornelius Rosse, José Mejino and Todd Detwiler, developers of the FMA at the University of Washington; to Alan Rector and Jeremy Rogers, developers of the GALEN at the University of Manchester, UK; to Martin Ringwald's group at the Jackson Laboratory for help with the Adult Mouse Anatomical Dictionary; and to Sherri de Coronado at

the National Cancer Institute for the NCI thesaurus.

Special thanks go to the researchers who collaborated with us closely on the evaluation of this study. Phil Bernstein at Microsoft Research and Peter Mork (then an intern in Phil's lab) shared with us the alignment they performed between the FMA and GALEN using a generic schema matching approach. Terry Hayamizu from the Jackson Laboratory established the manual alignment between NCI and MA used for evaluating our automatic mapping.

REFERENCES

Aleksovski, Z., Klein, M., ten Kate, W., & van Harmelen, F. (2006). Matching unstructured vocabularies using a background ontology. In S. Staab & V. Svatek (Eds.), *Managing knowledge in a world of networks -- Proceedings of the 15th International Conference on Knowledge Engineering and Knowledge Management (EKAW'06) -- LNAI 4248* (pp. 182-197). Berlin / Heidelberg: Springer

Aleksovski, Z., ten Kate, W., & van Harmelen, F. (2006). Exploiting the structure of background knowledge used in ontology matching. *Proceedings of the International Workshop on Ontology Matching (OM 2006), November 5, 2006, Athens, Georgia, USA*, 13-24.

Bard, J. B. (2005). Anatomics: the intersection of anatomy and bioinformatics. *J Anat, 206*(1), 1-16.

Bodenreider, O. (2004). The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res, 32 Database issue*, D267-270.

Bodenreider, O., Hayamizu, T. F., Ringwald, M., de Coronado, S., & Zhang, S. (2005). Of mice and men: Aligning mouse and human anatomies. *Proc AMIA Symp*, 61-65.

Bodenreider, O., & Zhang, S. (2006). Comparing the representation of anatomy in the FMA and SNOMED CT. *Proc AMIA Symp*, 46-50.

Burgun, A. (2005). Desiderata for domain reference ontologies in biomedicine. *J Biomed Inform*.

De Coronado, S., Haber, M. W., Sioutos, N., Tuttle, M. S., & Wright, L. W. (2004). NCI Thesaurus: Using Science-based Terminology to Integrate Cancer Research Results. *Medinfo, 2004*, 33-37.

Doan, A., & Halevy, A. Y. (2005). Semantic integration research in the database community: A brief survey. *AI Magazine, 26*(1), 83-94.

Doan, A., Madhavan, J., Domingos, P., & Halevy, A. Y. (2004). Ontology matching: A machine learning experience. In S. Staab & R. Studer (Eds.), *Handbook on Ontologies* (pp. 385-403): Springer-Verlag

Euzenat, J., Stuckenschmidt, H., & Yatskevich, M. (2005). Introduction to the ontology alignment evaluation 2005. In B. Ashpole, M. Ehrig, J. Euzenat & H. Stuckenschmidt (Eds.), *Proceedings of the K-CAP 2005 Workshop on Integrating Ontologies* http://ceur-ws.org/Vol-156/.

Goble, C. A., Stevens, R., Ng, G., Bechhofer, S., Paton, N. W., Baker, P. G., et al. (2001). Transparent Access to Multiple Bioinformatics Information Sources. *IBM Systems Journal Special issue on deep computing for the life sciences, 40*(2), 532-552.

Guarino, N. (1998a). Formal ontology in information systems. *Proceedings of FOIS'98*, 3-15.

Guarino, N. (1998b). Some ontological principles for designing upper-level lexical resources. *Proceedings of the First International Conference on Language resources and Evaluation*, 527-534.

Hayamizu, T. F., Mangan, M., Corradi, J. P., Kadin, J. A., & Ringwald, M. (2005). The Adult Mouse Anatomical Dictionary: a tool for annotating and integrating data. *Genome Biology, 6*(3), R29.

Horrocks, I., & Sattler, U. (2001). Ontology reasoning in the SHOQ(D) description logic. *Proceedings of the 17th Int. Joint Conf. on Artificial Intelligence (IJCAI 2001)*, 199-204.

Jian, N., Hu, W., Cheng, G., & Qu, Y. (2005). FalconAO: Aligning ontologies with Falcon. In B. Ashpole, M. Ehrig, J. Euzenat & H. Stuckenschmidt (Eds.), *Proceedings of the K-CAP 2005 Workshop on Integrating Ontologies* http://ceur-ws.org/Vol-156/.

Kalfoglou, Y., & Hu, B. (2005). CROSI Mapping System (CMS) - Result of the 2005 Ontology Alignment Contest. In B. Ashpole, M. Ehrig, J. Euzenat & H. Stuckenschmidt (Eds.), *Proceedings of the K-CAP 2005 Workshop on Integrating Ontologies* http://ceur-ws.org/Vol-156/.

Kalfoglou, Y., & Schorlemmer, M. (2002). Information-flow-based ontology mapping. In *Proceedings of the 1st International Conference on Ontologies, Databases and Application of Semantics (ODBASE'02)* (Vol. 2519, pp. 1132-1151)

Kalfoglou, Y., & Schorlemmer, M. (2003). Ontology mapping: the state of the art. *Knowledge Engineering Review, 18*(1), 1-31.

Kiryakov, A., Popov, B., Ognyanoff, D., Manov, D., Kirilov, A., & Goranov, M. (2003). Semantic annotation, indexing, and retrieval. In *Proceedings of the International Semantic Web Conference (ISCW) 2003* (Vol. 2870, pp. 484-499)

Klein, M. (2001). Combining and relating ontologies: an analysis of problems and solutions. *Proceedings of the IJCAI-2001 Workshop on Ontologies and Information Sharing*, 53-62.

Kotis, K., Vouros, G. A., & Stergiou, K. (2006). Towards automatic merging of domain ontologies: The HCONE-merge approach. *Web Semantics: Science, Services and Agents on the World Wide Web, 4*(1), 60-79.

Madhavan, J., Bernstein, P. A., & Rahm, E. (2001). Generic schema matching using Cupid. *Proceedings of 27th International Conference on Very Large Data Bases*, 49–58.

McCray, A. T., Srinivasan, S., & Browne, A. C. (1994). Lexical methods for managing variation in biomedical terminologies. *Proc Annu Symp Comput Appl Med Care*, 235-239.

Mork, P., & Bernstein, P. A. (2004). Adapting a generic match algorithm to align ontologies of human anatomy. *Proceedings of the 20th International Conference on Data Engineering*, 787-790.

Mork, P., Pottinger, R., & Bernstein, P. A. (2004). Challenges in precisely aligning models of human anatomy using generic schema matching. *Medinfo, 2004*, 401-405.

Noy, N. F. (2004a). Semantic integration: a survey of ontology-based approaches. *SIGMOD Rec., 33*(4), 65-70.

Noy, N. F. (2004b). Tools for mapping and merging ontologies. In S. Staab & R. Studer (Eds.), *Handbook on Ontologies* (pp. 365-384): Springer-Verlag

Noy, N. F., & Musen, M. A. (2000). PROMPT: algorithm and tool for automated ontology merging and alignment. *Proceedings of AAAI*, 450-455.

Noy, N. F., Musen, M. A., Mejino, J. J. L. V., & Rosse, C. (2004). Pushing the envelope: challenges in a frame-based representation of human anatomy. *Data & Knowledge Engineering, 48*(3), 335-359.

Rahm, E., & Bernstein, P. A. (2001). A survey of approaches to automatic schema matching. *VLDB Journal, 10*, 334-350.

Rector, A. L., Bechhofer, S., Goble, C. A., Horrocks, I., Nowlan, W. A., & Solomon, W. D. (1997). The GRAIL concept modelling language for medical terminology. *Artif Intell Med, 9*(2), 139-171.

Rogers, J., & Rector, A. (2000). GALEN's model of parts and wholes: experience and comparisons. *Proc AMIA Symp*, 714-718.

Rosse, C., & Mejino, J. L., Jr. (2003). A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform, 36*(6), 478-500.

Schulz, S., & Hahn, U. (2005). Part-whole representation and reasoning in formal biomedical ontologies. *Artif Intell Med, 34*(3), 179-200.

Shvaiko, P., & Euzenat, J. (2005). A survey of schema-based matching approaches. *Journal on Data Semantics, 4*, 146-171.

Smith, B., Ceusters, W., Klagges, B., Kohler, J., Kumar, A., Lomax, J., et al. (2005). Relations in biomedical ontologies. *Genome Biol, 6*(5), R46.

Stumme, G., & Maedche, A. (2001). FCA-Merge: Bottom-up merging of ontologies. *Prceedings of the 7th International Conference on Artificial Intelligence (IJCAI '01)*, 225-230.

Travillian, R. S., Gennari, J. H., & Shapiro, L. G. (2005). Of mice and men: Design a comparative anatomy information system. *AMIA Annu Symp Proc*, 734-748.

Winston, M., Chaffin, R., & Herrmann, D. (1987). A Taxonomy of part-whole relations. *Cognitive Science*(11), 417-444.

Zhang, S., & Bodenreider, O. (2003). Aligning representations of anatomy using lexical and structural methods. *Proc AMIA Symp*, 753-757.

Zhang, S., & Bodenreider, O. (2004a). Comparing associative relationships among equivalent concepts across ontologies. *Medinfo*, 459-463.

Zhang, S., & Bodenreider, O. (2004b). Investigating implicit knowledge in ontologies with application to the anatomical domain. In R. B. Altman, A. K. Dunker, L. Hunter, T. A. Jung & T. E. Klein (Eds.), *Pacific Symposium on Biocomputing 2004* (pp. 250-261): World Scientific

Zhang, S., & Bodenreider, O. (2005). Alignment of multiple ontologies of anatomy: Deriving indirect mappings from direct mappings to a reference. *Proc AMIA Symp*, 864-868.

Zhang, S., & Bodenreider, O. (2006a). Aligning anatomical ontologies: The role of complex structural rules. *Proceedings of the 18th International Conference on Systems Research, Informatics and Cybernetics (InterSymp 2006)*, (in press).

Zhang, S., & Bodenreider, O. (2006b). Law and order: Assessing and enforcing compliance with ontological modeling principles. *Computers in Biology and Medicine, 36*(6-7), 674-693.

Zhang, S., Bodenreider, O., & Golbreich, C. (2006). Experience in reasoning with the Foundational Model of Anatomy in OWL DL. In R. B. Altman, A. K. Dunker, L. Hunter, T. A. Murray & T. E. Klein (Eds.), *Pacific Symposium on Biocomputing 2006* (pp. 200-211): World Scientific

Zhang, S., Mork, P., & Bodenreider, O. (2004). Lessons learned from aligning two representations of anatomy. In U. Hahn, S. Schulz & R. Cornet (Eds.), *Proceedings of the First International Workshop on Formal Biomedical Knowledge Representation (KR-MED 2004)* (pp. 102-108)