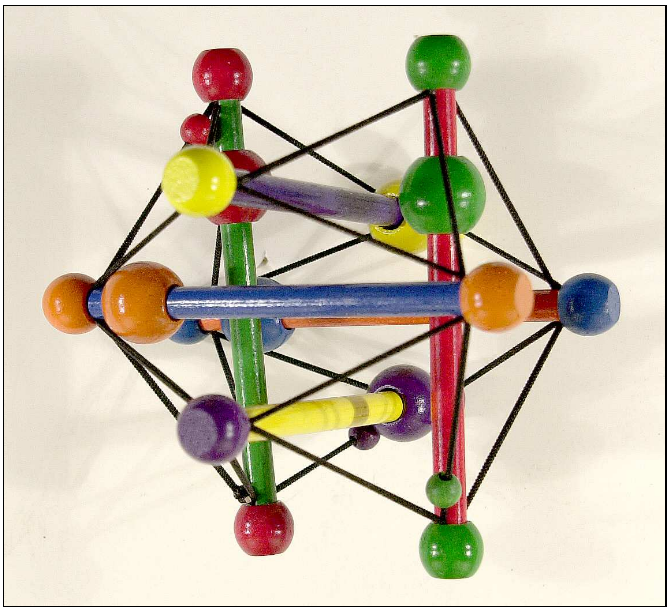
Medical Ontology Research

A Report to the Board of Scientific Counselors
of the Lister Hill National Center
for Biomedical Communications

May 17, 2001



Olivier Bodenreider

Table of Contents

1 Introduction

Ontology is classically defined as “a branch of metaphysics relating to the nature and relations of being”. According to Sowa, an ontology is a catalog of the types of things that are assumed to exist in a domain of interest. The ontology is designed from the perspective of a person who uses a language for the purpose of talking about the domain. The types in the ontology represent the predicates, word senses, or concept and relation types of the language when used to discuss topics in the domain [1]. Sowa adds: “the combination of logic with an ontology provides a language that can express relationships about the entities in the domain of interest”. Gruber, more illustratively, sees an ontology, as “a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents” [2]. He also defines ontological commitment as “an agreement to use a vocabulary (i.e., ask queries and make assertions) in a way that is consistent (but not complete) with respect to the theory specified by an ontology”.

In the biomedical domain, defining and organizing knowledge has been a concern for several centuries. The system invented by Carolus Linnaeus in the mid-18th century for naming, ranking, and classifying living organisms is still in use today, although modified. Another example is given by Thomas Sydenham, whose work created a framework for the classification of diseases. By observing clinical phenomena at the patient’s bedside, he conceived for the first time the possibility of a variety of distinct diseases, as opposed to a general illness caused by the “imbalance of humours”.

While existing representations of the biomedical domain may be sufficient for information retrieval purposes, the organization of knowledge in these representations is generally not suitable for reasoning. In fact, reasoning requires the principled, consistent and organization usually provided by ontologies.

The objective of the MEDICAL ONTOLOGY RESEARCH project is *not* to build an ontology of the biomedical domain. Rather, our goal is to develop methods whereby ontologies could be acquired from existing resources, as well as validated against other knowledge sources. To begin with, however, we think that shifting the approach from ontologies to semantic spaces might help make progress toward knowledge organization while making it possible to articulate this research project with other projects or applications developed at NLM. We propose to investigate several aspects of semantic spaces in the medical domain including definition, organization, visualization, and utilization.

2 Background

In this section, we present the differences between terminologies, semantic spaces and ontologies. We then present the Unified Medical Language System® (UMLS®). Finally, we examine some aspects of knowledge organization in the components of the UMLS.

2.1 Terminologies, semantic spaces and ontologies

Terminologies, semantic spaces and ontologies, although distinct, may be thought of as a several levels in knowledge organization, each serving its own purpose.

Terminology is concerned by the naming problem rather than by the organization of knowledge. Medical terminologies provide a controlled set of language labels, or terms, attached to a concept, or meaning, in the biomedical domain. While some medical terminologies are flat lists of terms, also called nomenclatures, others bear some kind of organization (single-inheritance or polyhierarchical structures, multi-axial system) [3]. When existing, knowledge organization in medical terminologies is not necessarily principled or consistent. For example, while the relation used to define hierarchies is either taxonomic (*isa*) or meronomic (*part of*), most vocabularies allow, often implicitly, other relations to be used for creating their hierarchical structure (e.g., *manifestation of*). Although such hierarchical structures may prove useful for information retrieval, they do not support inheritance and thus do not allow reasoning.

With semantic spaces, the objective is essentially to represent the relationships among concepts. While some relationships are inherited from the structure of the terminologies (e.g., hierarchies), other relationships can be acquired using Natural Language Processing techniques (lexical knowledge, based on some degree of resemblance between terms) or statistical techniques (statistical knowledge, based on the co-occurrence of concepts in a corpus). In semantic spaces, although the clusters of related concepts that obtain may be of interest in some applications (e.g., information retrieval), the semantics of the inter-concept relationships acquired through various techniques may not be defined with precision (e.g., co-occurring concepts). Additional constraints are needed for a sound knowledge representation in order to support, for example, reasoning.

According to Sowa, in practice, a formal ontology "is specified by a collection of names for concept and relation types organized in partial ordering by the type-subtype relation" [1]. These constraints clearly represent a level of organization that is not expected from semantic spaces, even if a semantic space contains all the concepts needed for representing a given domain.

Rather than completely distinct structures, terminologies, semantic spaces and ontologies may be thought of as a sort of continuum in knowledge organization, from less organized (terminologies) to more organized (ontologies). Thus, building an ontology does not necessarily mean creating it from scratch, but may rather be achieved by adding formality and consistency to the organization of a partially structured set of concept (e.g., a semantic space).

2.2 The Unified Medical Language System

As a source of biomedical knowledge, we will use in particular the Unified Medical Language System (UMLS), developed and maintained by the National Library of Medicine since 1990. The UMLS is intended to help health professionals and researchers use biomedical information from different sources [4]. The UMLS comprises two major inter-related components: the Metathesaurus®, a huge repository of concepts, and the Semantic Network, a limited network of semantic types [5].

The current version (2001) of the Metathesaurus integrates about 800,000 concepts from more than fifty families of vocabularies such as the International Classification of Diseases or Medical Subject Headings. While the structure of each source vocabulary is preserved, terms that are equivalent in meaning are clustered into a unique concept. Furthermore, inter-concept relationships, either inherited from the source vocabularies or specifically generated, give the UMLS Metathesaurus additional semantic structure. The UMLS building process imposes no restrictions on the source vocabularies prior to integrating their terms and structure into the Metathesaurus. Therefore, hierarchical relationships in the Metathesaurus are not expected to represent homogeneous taxonomic relations, but rather to reflect the several organizational principles inherited from the source vocabularies.

The UMLS Semantic Network is a network of 134 semantic types used to categorize Metathesaurus concepts. A definition is given for each semantic type. The semantic types are organized in two high-level single-inheritance hierarchies, one for entities, one for events. The *isa* link allows nodes to inherit properties from higher-level nodes. In addition, associative relationships divided into five subcategories (physical, spatial, functional, temporal, conceptual relationships) are instantiated between the semantic types. They represent general high-level knowledge, such as “drugs treat diseases”. Conversely, Metathesaurus inter-concept relationships instantiate specific low-level knowledge, such as “aspirin treats fever”. When two semantic types are linked by some relationship, the relationship may hold or not for any particular pair of concepts that have been assigned to those semantic types (obviously, not every drug treats every disease).

Each Metathesaurus concept is assigned to at least one semantic type from the Semantic Network, providing each concept a categorization that is independent from its relationships to other concepts.

2.3 Toward an ontology of the biomedical domain

The organization of knowledge using a two-level structure is specific to the UMLS. The Semantic Network defines and organizes a small number of semantic types. Assuming that it is compatible with general upper-level ontologies and that all needed categories are present, the Semantic Network, represented as a type hierarchy, may be thought of as an upper-level ontology of the biomedical domain.

The organization of the Metathesaurus, on the other hand, relies on different principles. Since the UMLS does not censor any information provided by the source vocabularies, hierarchical relationships sometimes reflect the several organizational principles inherited from the source vocabularies and may fail to have the properties of partial ordering relations. For this reason, the Metathesaurus fails to meet basic ontological requirements.

The categorization of Metathesaurus concepts with semantic types realizes the link between the two structures. This link allows Metathesaurus concepts for inheriting properties defined for types at the level of the Semantic Network. Even though, the additional organizational structure brought to the Metathesaurus by the Semantic Network is not sufficient to turn the Metathesaurus into an ontology of the biomedical domain.

However, assuming that it is possible to further refine its organizational structure, the Metathesaurus can provide the foundation for an ontology of the biomedical domain. While semantic spaces are useful representations for information retrieval, ontologies are needed for tasks such as natural language processing and reasoning.

As we said earlier, developing methods whereby ontologies could be acquired from existing resources is the ultimate goal of the MEDICAL ONTOLOGY RESEARCH project. An intermediate, more achievable goal is to fully exploit the UMLS as a semantic space. Later, the organizational structure of the semantic space will be refined in order to turn it into an ontology. This project is expected to require basic research (e.g., to study the specificity of the taxonomic relation in the biomedical domain) as well as more applied research (e.g., to develop UMLS-based algorithms used in applications). A sound, ontological representation of biomedical knowledge is expected to enable tasks such as reasoning.

The following four sections present aspects of this research organized into four major themes: Definition, organization, visualization and utilization of semantic spaces in the medical domain.

3 DEFINE

Knowledge associated with a concept, i.e. its definition and its relationships with other concepts refers to both symbolic representation and statistical information, from which semantic spaces can be constructed [6]. Semantic spaces can be defined from semantic information *provided* by existing terminologies (symbolic knowledge represented by relationships among concepts), knowledge bases, and expert systems (e.g., rules and facts). Additional information useful for defining semantic spaces can also be *extracted* from the medical literature using, for example, natural language processing techniques (lexical knowledge). Finally, statistical knowledge acquired, for example, from the co-occurrence of biomedical concepts in large corpora of text represents a valuable source of knowledge, in complement to symbolic knowledge. Figure 1 shows an example of such a semantic space surrounding the concept “Heart”, built from symbolic and statistical knowledge available in the UMLS.

We studied some formal aspects of the relationships used for representing knowledge in the biomedical domain as a prerequisite to defining semantic spaces. We have proposed methods for studying the semantics of the relationships between co-occurring concepts. Knowledge acquisition techniques (e.g., using lexical knowledge for automatic acquisition of hyponymic relations) are useful for both defining and organizing semantic spaces, and some techniques will be presented in the next section (4). In the future, we plan to explore methods for defining subspaces and to develop a measurement of the semantic distance between biomedical concepts in the UMLS.

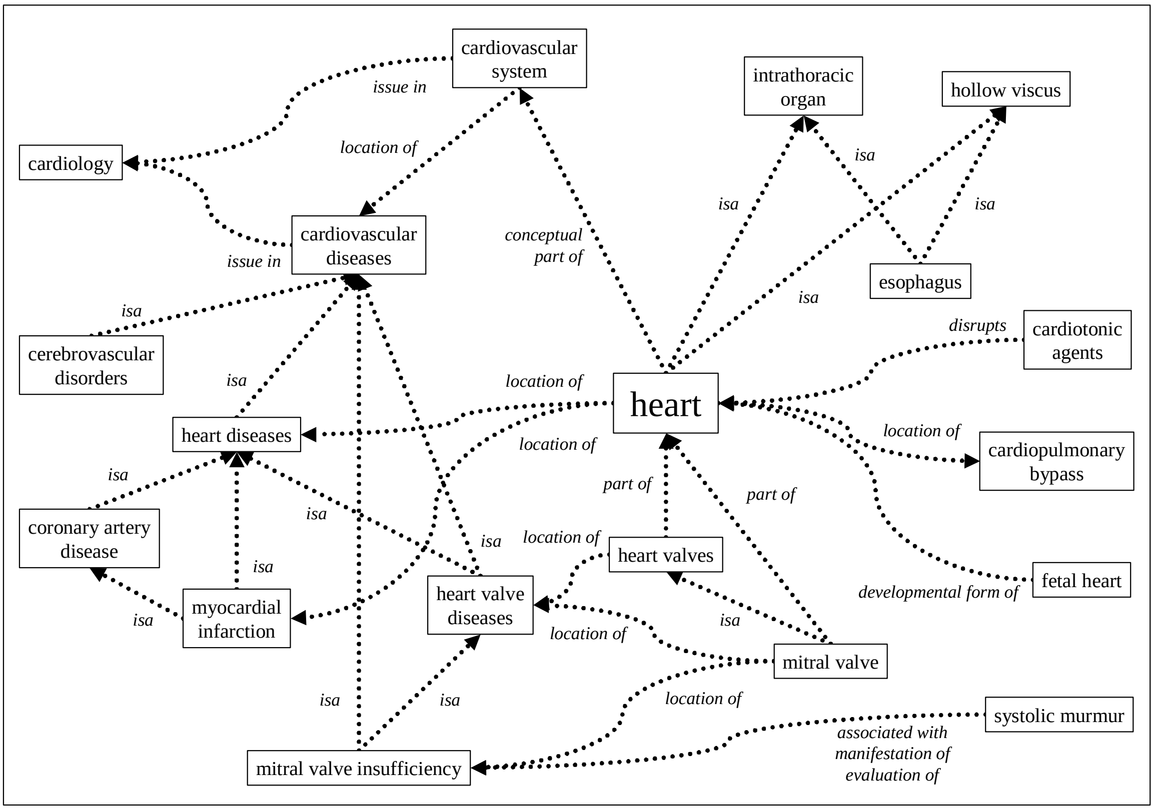


Figure 1 - Semantic space around the concept "Heart" in the UMLS (partial representation)

3.1 Formal aspects of the relationships

The identification and augmentation of relationships among knowledge structures inherited from medical vocabularies has been an early goal in the UMLS project and has been recognized as an important part of the added value in the Metathesaurus. So, quite naturally, we used the UMLS to study and illustrate the role played by relationships in the integration of biomedical vocabularies, in the broader context of a book about relationships in the organization of knowledge [7]. We focused on three types of relationships among terms: synonymy, hierarchical relationships and explicit mapping relationships, showing how underlying knowledge structures are connected through these relationships. For each kind of relationships, we presented some formal aspects, we examined how the relationship is implemented in the UMLS, and we investigated some of the issues raised by this relationship in the biomedical domain. Synonymy-related issues include the presence of underspecified terms (e.g., “Hodgkin’s disease, not otherwise specified") and classification-specific terms (e.g., "Other testicular dysfunction") as well as the use of implicit knowledge from the context of a term (e.g., "prostate" for "prostate cancer" when used in a section about "cancers of genital organs"). The presence of circular hierarchical relationships is one of the issues related to hierarchies.

In a more recent study, we focused on the taxonomic relation in the biomedical domain [8]. Taxonomies are commonly used for organizing knowledge, particularly in biomedicine where the taxonomy of living organisms and the classification of diseases are central to the domain. The principles used to produce taxonomies are either intrinsic (properties of the partial ordering relation) or added to make knowledge more manageable (opposition of siblings and economy). The applicability of these principles in the biomedical domain was presented using the UMLS and issues raised by the application of these principles were illustrated. While intrinsic principles were not challenged, we argued that the opposition of siblings brings to bear excessive constraints on a domain ontology and that the adverse effects of economy may outweigh its benefits.

3.2 Semantics of the relationships between co-occurring concepts

In order to study how statistical knowledge based on co-occurring concepts contributes to the definition of semantic spaces, we analyzed the co-occurrence of MeSH descriptors in MEDLINE citations (1990-1999). 18,485 UMLS concepts involved in 7,928,608 directed pairs of co-occurring concepts were studied. For each directed pair of concepts C_1-C_2: (i) the "family[1]" of C_1 was built, using the UMLS Metathesaurus, and we tested whether or not C_2 belonged to C_1's family; (ii) we used the semantic categorization of Metathesaurus concepts through the UMLS Semantic Network and Semantic Groups to represent the semantics of the relationships between C_1 and C_2.

In 6.5% of the directed pairs, the co-occurring concept C_2 was found within the "family" of C_1. This means that most of the semantics of the relationships between co-occurring concepts is not redundant with the symbolic knowledge recorded in the UMLS. In terms of semantic groups, the most frequent association is found between two concepts of the group "Chemicals & Drugs"; another frequent pattern associates concepts from the groups "Disorders" and "Chemicals & Drugs". The semantics of the relationships between co-occurring concepts can be partially inferred from the pairs of semantic groups. For example, the relationship between "Disorders" and "Anatomy" should be mainly "has location". Nevertheless the relationship between "Disorders" and "Chemicals & Drugs" remains ambiguous, since it could be either "treated by" or "caused by".

This study will be presented at MEDINFO'2001 [9].

[1] In addition to the relationships available in the UMLS, we **redefined hierarchical relationships** by combining, for example, parents and broader concepts into "First-generation Ancestors", and hierarchically-related relationships by extending the notion of siblings to children ot narrower concepts of parents or broader concepts. We also defined **multiple-level relationships** (e.g., Ancestors or Descendants, all the way to the top or to the bottom of hierarchies), and **combined relationships** (e.g., uncles, the extended siblings of first-generation ancestors, and cousins, the first-generation descendants of uncles).

3.3 Research directions

As a task, defining **semantic subspaces** is useful for extracting concepts (and thus vocabulary) corresponding to a subdomain according to an organ or a body system (e.g., cardiology), or to a given procedure (e.g. transplantation). The subspaces that obtain can then be combined by means of operations such as intersection or union in order to define new semantic spaces. In addition to extracting the vocabulary for a subdomain, applications of semantic spaces restricted to a subdomain include the disambiguation of polysemous terms, by comparing their context to several candidate semantic spaces.

Calculating a **semantic distance** between two concepts is a difficult task because, unlike vector spaces in algebra used, for example, in multi-dimensional data analysis of numeric data, semantic axes have no built-in scale and there are no absolute metrics applicable to semantic relations. However, several models have been proposed to define a distance, or, at least, a measure of proximity, closeness or similarity [see, for example, 10, 11]. While some models rely essentially on the hierarchical structure of a thesaurus [e.g., 12], others take advantage of statistical knowledge [e.g., 13]. In the UMLS, we plan to explore a definition of semantic distance combining symbolic and statistical knowledge. It is expected that the semantic distance will provide a quantitative measure for semantic locality, and thus will be used as a simpler way to define semantic spaces.

4 ORGANIZE

Organizing biomedical knowledge is one of the goals of the UMLS. However, the current level of organization is not consistent and principled enough to fully support reasoning. One central aspect of this project is therefore to investigate how knowledge representation would need to be improved to fulfill ontological requirements.

We have already investigated some problems existing in the UMLS and often reported them in the literature (e.g., circular hierarchical relationships and discrepancies between Metathesaurus and Semantic Network relationships). We also compared the UMLS to ontologies such as Cyc or WordNet. We explored some alternative methods for acquiring knowledge. In the future, we plan to study ontological issues such as “what needs to be represented?”, to explore formalisms such as description logics for knowledge representation, to further investigate alternative methods for acquiring knowledge, and to compare the UMLS to other modern systems (e.g., *Open*GALEN and SNOMED-RT).

4.1 Analyze existing problems

Virtually any evaluative study of the Metathesaurus with a focus on semantics mentions problems such as the presence of circular hierarchical relationships, inconsistencies in the categorization of the concepts, and discrepancies between the semantic structure of the Semantic Network and that of the Metathesaurus [e.g., 14, 15]. We recently analyzed the circular hierarchical relationships in the Metathesaurus. We also compared inter-concept relationships in the Metathesaurus to the relationships between their corresponding semantic types in the Semantic Network on a limited subset of concepts. In both cases, we suggested that these analytical methods based on semantic principles be used in the system that helps build the Metathesaurus, either by implementing the semantic principles into this environment, or by systematically presenting UMLS editors with cases that violate those principles.

4.1.1 Circular hierarchical relationships in the UMLS

The UMLS building process imposes no restrictions on the source vocabularies prior to integrating their terms and structure into the Metathesaurus. In the source vocabularies, hierarchical relationships are usually not limited to taxonomic relations, but rather reflect the way each vocabulary organizes its terms, according to its purposes. Moreover, the precise nature of the relationship is mentioned in only about 25% of the cases; and, because many non-taxonomic relations are used to build hierarchies, it is not possible to assume that a non-labeled hierarchical relationship is probably taxonomic.

Even though they are heterogeneous, the organizational principles used to create hierarchies are expected to share some fundamental characteristics, and, thus, to be compatible. The mathematical relation associated with hierarchies is a *partial* ordering relation, which means that it is possible for a concept to be hierarchically related to itself (reflexive relation). However, in a directed acyclic graph, i.e. the data structure resulting from polyhierarchy, no path is allowed to start and end at the same vertex, which means that, when represented in a graph, the reflexive hierarchical relationships create cycles of a particular kind, called loops. In practice, we make no distinction among circular hierarchical relationships on the basis of the number of concepts involved in the cycles, because any cycle has similar detrimental consequences in terms of graph traversal.

Causes for circular hierarchical relationships in the Metathesaurus include:

- **Granularity**: terms in hierarchical relationship in a source vocabulary are considered synonymous in the Metathesaurus, and thus clustered into the same concept, which creates a reflexive relationship (Figure 2).

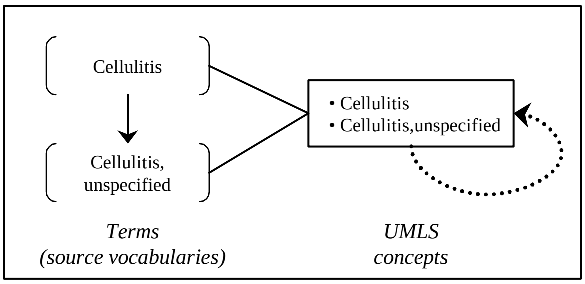


Figure 2 - Reflexive relationship in the Metathesaurus

- **Unspecified terms**: 62,000 UMLS terms bear some kind of underspecification marker, the most frequent being “unspecified” and “not otherwise specified”. In most source vocabularies, "T, unspecified" is a descendant of "T", while in the Metathesaurus, since they present no difference in meaning, they are generally clustered into the same concept, which creates a circular relationship[2] (Figure 3).

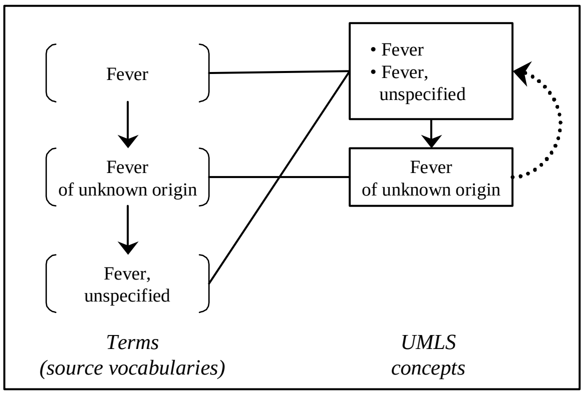


Figure 3 - Direct circular hierarchical relationship in the Metathesaurus

- **Classes, instances, and implicit data**: Although in most instances, inflectional variation of terms does not modify the meaning, in some cases, however, the plural form refers to a class, while the singular form refers to an instance, but not necessarily of the same class. For example "purine" is a heterocyclic compound that contributes to produce "purines" (the purine bases). There are two distinct concepts in the UMLS for "purine" and "purines". On the other hand, the terms "Topographic regions" and "body region" are considered synonymous in the UMLS, although "Topographic regions" in the particular context of SNOMED International actually groups a whole range of physical anatomical entities, including "body regions" (Figure 4). The implicit knowledge associated with a term used in a particular context is difficult to detect and is often not recognized.

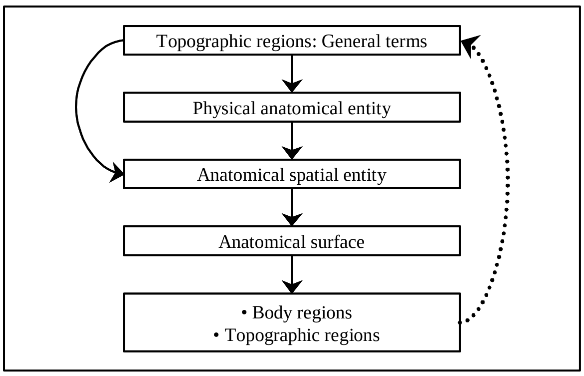


Figure 4 - Indirect circular hierarchical relationship in the Metathesaurus

[2] The circular relationship is reflexive when "T, unspecified" is a direct descendant (child) of "T"

Other causes for circular hierarchical relationships in the Metathesaurus include the presence of metadata in terms (e.g., “HEART DISEASES” and “HEART DISEASES: GENERAL TERMS”), the ambiguity inherent to compound terms in many terminologies (“nausea and vomiting” being used to designate either the association of “nausea” to “vomiting”, or “nausea or vomiting”), and organizational conventions.

This analysis was recently submitted for presentation at the AMIA Annual Fall Symposium in 2001 [16].

4.1.2 Discrepancies between Metathesaurus and Semantic Network relationships

In order to illustrate the ability and the limits of the UMLS Semantic Network to provide a conceptual framework for the biomedical domain, we designed the following experiment. Starting from a given concept, “heart”, we gathered the 3764 concepts that constitute its semantic neighborhood by exploiting a set of inter-concept relationships represented in the UMLS Metathesaurus. For each pair of related concepts from this set, we calculated the possible relationships between the concepts using the semantic links defined in the UMLS Semantic Network between the semantic types that had been assigned to these concepts. Besides revealing the semantic structure in this set of concepts, other expected results included qualifying broadly defined relationships in the Metathesaurus, assessing already defined ones, and, more generally, by enforcing semantic rules, detecting inconsistencies in the Metathesaurus or in the Semantic Network itself.

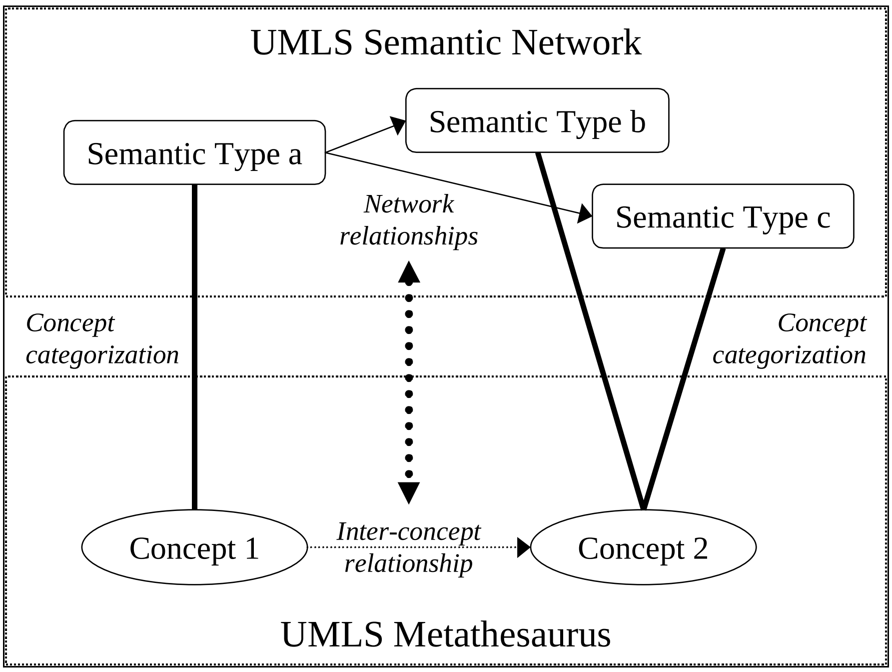


Figure 5 - Comparing Metathesaurus and Semantic Network relationships

The top part of Figure 5 represents the semantic types and the relationships between them, as defined in the Semantic Network. The Metathesaurus, a set of concepts linked by inter-concept relationships, is represented in the bottom part of the figure. The two structures are related by means of the semantic types assigned to the concepts in order to categorize them. Therefore, inter-concept relationships can be inferred, validated or rejected by comparison to the relationships defined between the semantic types assigned to the concepts.

Among the 6894 pairs of related concepts, we obtain the following results:

- In 4496 cases (65%), a semantic relation can be inferred unambiguously from the Semantic Network. The semantic relation inferred allows us to determine inter-concept relationships whose attribute was not defined in 2515 of these cases, and to confirm the validity of the relationship attribute in 1981 of these cases.
- In 1491 cases (22%), multiple semantic links exist between the semantic types of the two concepts, leading to several possible attributes for these inter-concept relationships.
- In the remaining 907 pairs (13%), the inter-concept relationships represent a violation of the Semantic Network. In 372 pairs, there is no semantic link between the semantic types of the two concepts. In 415 pairs, the inter-concept relationship is not compatible with that of the corresponding Semantic Network relationship. Finally, in 120 pairs, the attribute of the inter-concept relationship is not compatible with the semantic relationships allowed between the semantic types of the two concepts.

This method helps detect discrepancies between the semantics of the Metathesaurus and the semantics expressed by the Semantic Network. Automatic detection helps limit the need for human review by focusing on conflicting relationships that violate the semantic rules. One major cause for such a violation is somewhat artificial: concepts with an abstract semantic type (e.g. “Classification”) may have related concepts having a concrete semantic type (e.g. “Body Part, Organ, or Organ Component”). These types would be unrelated in the Semantic Network. The relationship of “heart: general terms” to “right side of heart”, for example, violates the Semantic Network for this reason. Another source of problems is that partonymic relationships (*part of*) are considered associative in the Semantic Network, while in many medical vocabularies they are used hierarchically. Frequently occurring semantic discrepancies may also help identify missing semantic links in the Semantic Network. For example, the relationship of “chest pain” to “thorax” violates the Semantic Network, since the *location of* relationship has not been defined between a “Body Location or Region” and a “Sign and Symptom”.

This study is part of the book “Semantics of Relationships” to be published in 2001 [17].

4.2 Compare to existing ontologies

The UMLS is sometimes presented as an ontology of the biomedical domain. Such an expectation is far beyond what the UMLS has been designed for. Actually, as we mentioned earlier, although it provides a representation of the biomedical domain suitable for some applications such as information retrieval, the UMLS does not necessarily fulfill ontological requirements. Being created from the bottom-up, i.e. integrating existing biomedical vocabularies without imposing any restrictions, the UMLS cannot *enforce* ontological principles.

Nevertheless, the UMLS having the potential of providing the basis for an ontology of the biomedical domain, we explored its compatibility with other ontologies.

We selected two widely distributed ontologies for they represent, arguably, two major aspects of what an ontology is expected to provide: Upper Cyc Ontology and WordNet. The choice of Cyc® was motivated by the fact that it provides a sufficient general grounding, while it may encompass microtheories. The structure of WordNet®, on the other hand, is closer to that of the Metathesaurus (terms, concepts, hierarchies), and WordNet captures common sense knowledge.

By investigating ontologies outside the biomedical domain, we would like also to reach out the medical informatics community, and to develop exchanges with communities involved in ontology and terminology.

4.2.1 Comparing terms, concepts and semantic classes in WordNet and the UMLS

We designed a study to compare how a general terminological system (WordNet) and a domain-specific one (UMLS) represent linguistic and knowledge phenomena at three different levels: terms, concepts, and semantic classes. For one general class (ANIMAL) and one domain-specific class (HEALTH DISORDER), the set of concepts corresponding to the class was established. Then, for each semantic class, the corresponding terms were mapped from one system to the other, both ways.

Only 2% of the domain-specific concepts from UMLS were found in WordNet, but 83% of the domain-specific concepts from WordNet were found in the UMLS. Concept overlap between the two systems varies from 48% to 97%.

This comparison reveals missing terms in both systems. Beside concepts specific to the medical domain, the UMLS contains specific terms for concepts found in both systems, including inverted terms (e.g., “Epilepsy, generalized”), and terminology-specific terms (e.g., “Generalized epilepsy, *without mention of intractable epilepsy*”). On the other hand, WordNet also has disease terms that are not found in the UMLS. These terms are generally lay synonyms (e.g., “kissing disease” for “infectious mononucleosis”). This phenomenon is of potential interest for augmenting lay terminology in the UMLS, with applications in consumer health projects, for example.

This study will be presented in June at the NAACL workshop “WordNet and other lexical resources: Applications, extensions, and customizations” [18].

4.2.2 Mapping the UMLS Semantic Network into General Ontologies

We analyzed the compatibility between the UMLS Semantic Types and two general ontologies: Cyc and WordNet.

Descriptions of the semantic types in the Cyc formalism were performed manually, using the hierarchical relationships available in Cyc. Additional Cyc categories were used as required to insure consistency. The relationship between a given semantic type T and the closest Cyc concept U was either *Similarity* if T had an equivalent U, or *Overlap* if there was a partial overlap between T and U (T and U being compatible and having a common supertype). One fifth of the UMLS semantic types had exact mapping to standard Upper Cyc Ontology. Despite its lack of depth in the biomedical domain, Cyc provides generic concepts and a structure that relies on more numerous, more richly organized categories,.

The mapping of UMLS and WordNet classes was based on comparing the sets of concepts that are subsumed by a given semantic type in the UMLS, and the sets of hyponyms of a given synset in WordNet. We focused on two classes: ANIMAL, which is a general class, supposed to be similarly represented in both systems, and HEALTH DISORDER, which is a typically medical class. 2% of the UMLS concepts from the HEALTH DISORDER class were present in WordNet, and compatibility between classes was 48%. WordNet, as a general language-oriented ontology is a source of common sense knowledge and lay terms, particularly important for consumer health applications.

This analysis was recently submitted for presentation at the AMIA Annual Fall Symposium in 2001 [19].

4.3 Explore alternative methods to acquire knowledge

Although the Metathesaurus comprises an extensive collection of inter-concept relationships, not all possible relationships are represented, and, when they are, the nature of the relationship (e.g., *isa*, *location of*) is explicitly mentioned in only about 25% of the cases. Therefore, knowledge acquisition from other sources or by techniques other than those used to build the Metathesaurus is expected to help discover missing relationships, as well as validate and qualify existing ones.

The comparison of relationships at two levels of the UMLS (in the Semantic Network and in the Metathesaurus) presented earlier (4.1.2) provides a proof of concept relying entirely on UMLS data. Statistical knowledge (e.g., the co-occurrence of MeSH descriptors in MEDLINE citations) can be seen as an alternate source of inter-concept relationships, whose semantics needs to be made explicit. We also explored the acquisition of hyponymic relations using lexical techniques (adjectival modification).

4.3.1 Lexically-suggested hyponymic relationships

Among the various methods for identifying thesaurus relations from text corpora, methods based on head modifier relation are interesting in the context of medical terminologies, especially for those terms which differ from one another by only one modifier. Adjectival modifiers play a particular role because they usually introduce a hyponymic relation. This study focused on comparing lexically-suggested hyponymic relations among medical terms to inter-concept relationships represented in the UMLS Metathesaurus.

A set of 63,000 medical terms from SNOMED International was selected, representing terms for diseases and procedures. Adjectival modifiers were identified from these terms, and transformed terms were generated by removing modifiers from the original terms. Candidate hyponymic relations were then tested against inter-concept relationships recorded in the UMLS Metathesaurus.

In 50% of the cases, suggested hyponymic relations were present in the UMLS Metathesaurus. In 25% of the cases, the original term and the transformed terms were "siblings" in the UMLS. In the remaining 25%, no relationship was recorded in the UMLS between these two terms.

The lack of relationships observed in the UMLS Metathesaurus was analyzed.

- **Lack of organization within a source vocabulary**. As an example of this phenomenon, although "acute infantile eczema" is a hyponym of the three terms "acute eczema", "infantile eczema" and "eczema", only the relationship to "disease of the skin and subcutaneous tissues", provided by SNOMED, is represented in the UMLS for "acute infantile eczema".
- **Lack of links across vocabularies**. Although the partially organized list of terms from SNOMED acquires an additional structure through relationships contributed by other source vocabularies or by the UMLS editors, specialized terms that appear only in one vocabulary (e.g. "acute infantile eczema") sometimes fail to be linked to some hypernym.
- **Underspecified terms**. The UMLS Metathesaurus provides several examples of confusion between the generic concept represented by a term T and the most frequent meaning of T. This phenomenon is extremely frequent in the biomedical domain, where numerous modifiers are implicit in medical terms. For example, "hip dislocation" and "acquired hip dislocation" are synonyms in the Metathesaurus while, in fact, hip dislocation may be either congenital or acquired by traumatism, even if the typical, most frequent form for hip dislocation is traumatic. As a result, "congenital hip dislocation" becomes a hyponym of "hip dislocation", while "acquired hip dislocation" is a synonym of "hip dislocation". In addition, "congenital hip dislocation" also becomes a hyponym of "acquired hip dislocation", which is incorrect.

Missing synonymy and the existence of micro-relations (close hyponyms clustered into the same concept synonyms in the Metathesaurus) also participate to the lack of relationships observed in the UMLS.

We proposed some methods for automatically assessing the suggested hyponymic relations are proposed. Alternatively, this method could be integrated in the Metathesaurus production environment: Lexically-suggested hyponymic relations could become candidate hierarchical relationships to be reviewed by the UMLS editors.

This study was recently presented at the Fourth Conference on "Terminology and Artificial Intelligence" in Nancy, France [20].

4.4 Research directions

As we mentioned earlier, the issue of organizing biomedical knowledge is central to this research project. As a general goal, we plan to develop methods for extracting from the UMLS a subset of terms, concepts and relationships compatible with ontological principles. This effort can be thought of as complementary, from an ontological perspective, to what we proposed recently for selecting useful strings from the UMLS, from a natural language processing perspective [21].

In practice, we plan to study what needs to be represented in an ontology of the biomedical domain. In the UMLS, some complex terms do not correspond to a single meaning but are nevertheless represented as concepts. These "concepts" must not be represented in an ontology. More generally, we plan to investigate if compositionality would constitute an alternative to term precoordination. A "core" Metathesaurus, comprising only all atomic concepts and the precoordinated concepts sanctioned by usage, is expected to result from this effort.

Another goal will be to validate, refine, and augment the relationships represented in the UMLS. Alternative sources of knowledge will be used for the validation of UMLS knowledge including other systems (e.g., *Open*GALEN and SNOMED-RT), other methods (e.g., statistical methods, lexical techniques, and semantic interpretation). Finally, we plan to explore formalisms such as description logics as a way to improve the representation of knowledge in the UMLS.

5 VISUALIZE

Making medical knowledge available to applications and making it accessible as well as navigable to users share some common issues. These issues include reducing complexity, providing consistent views across the domain, and extending the views provided to the domain in order to fit specific needs.

Several tools have been developed for accessing and visualizing UMLS data. MetaCard® has made it possible for users to browse the content of the Metathesaurus since its first version in 1990 [22]. Developed later, the Knowledge Source Server[3] (KSS) allows users to browse through a web-based interface, and manipulate through an application programming interface (API), virtually every bit of information in the UMLS [23]. In order to circumvent some of the limitations of KSS, we have developed an object-oriented model for manipulating UMLS knowledge as well as a browser for displaying and navigating this knowledge.

5.1 Object-oriented model for representing semantic locality

The representation of meaning in the UMLS allows users to define and explore the semantic space surrounding a given concept [24]. The various semantic links among concepts represent one of the organizing principles of the UMLS: semantic locality [25]. The dimensions of semantic locality include term information (synonymy, hypernymy, hyponymy), contextual information in a particular source, co-occurrence of terms in the medical literature, and the categorization of the concepts in a Semantic Network. Figure 6 shows a subset of the semantic space for the concept "Heart", based on the principles of semantic locality.

[3] umlsks.nlm.hih.gov

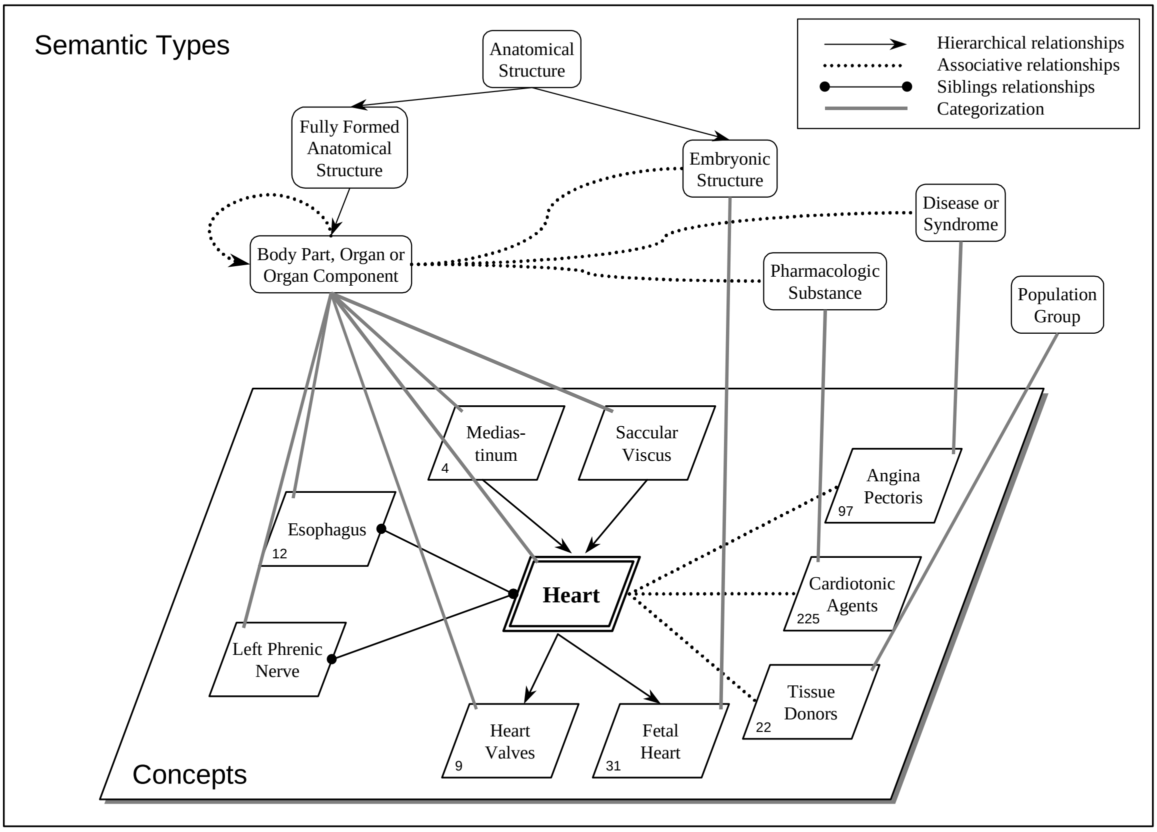


Figure 6 - Semantic space for the concept "Heart" (partial representation). Numbers refer to the frequency of co-occurrence in MEDLINE between "Heart" and other concepts, when available

The integration of UMLS data into an application remains difficult because KSS gives access to UMLS data through predefined queries but does not provide an information model. We developed an object-oriented model in which the semantic features of the UMLS are made available through four major classes for representing Metathesaurus concepts, semantic types, inter-concept relationships and Semantic Network relationships. Additional semantic methods for reducing the complexity of the hierarchical relationships represented in the UMLS are implemented, including methods for performing the transitive reduction of a graph of UMLS concepts. This model is used internally at NLM in a variety of applications, and will be presented at MEDINFO'2001 [26].

5.2 UMLS Semantic Navigator

In KSS, the presentation of contextual information from the Metathesaurus tends to reflect the organization of terms in the source vocabularies. For example, hierarchically related concepts are presented as lists of indented terms, one list for each vocabulary using this term. Instead of multiple trees, the semantic structure of the UMLS can be visualized as a graph in which concepts are the nodes and inter-concept relationships are the links between nodes. The graph structure offers a unified view of the context (Figure 6).

We developed an experimental semantic navigation tool – the UMLS Semantic Navigator – that is less comprehensive than KSS, but offers alternative display and navigation features, suitable for knowledge exploration. This tool accesses UMLS data through the object-oriented model presented in section 5.1. Through a web-based interface, the display puts a concept at the center of the screen and organizes related concepts on the periphery (Figure 7). Siblings, other related concepts, and co-occurring concepts are displayed as lists, while hierarchically-related concepts are represented graphically, using GraphViz[4], a graph visualization tool. Every concept presented on the screen is a hyperlink, so that a click on the concept moves it to the center of the screen and displays the related concepts corresponding to its semantic locality. Lists of related concepts are organized by semantic groups [27]. Several editions of the UMLS are available online. Additionally, a transitive reduction can be performed on the graph of hierarchically-related concepts in order to reduce the conceptual complexity. The UMLS Semantic Navigator is available to UMLS licensees[5], and was presented at the AMIA Annual Fall Symposium in 2000 [28].

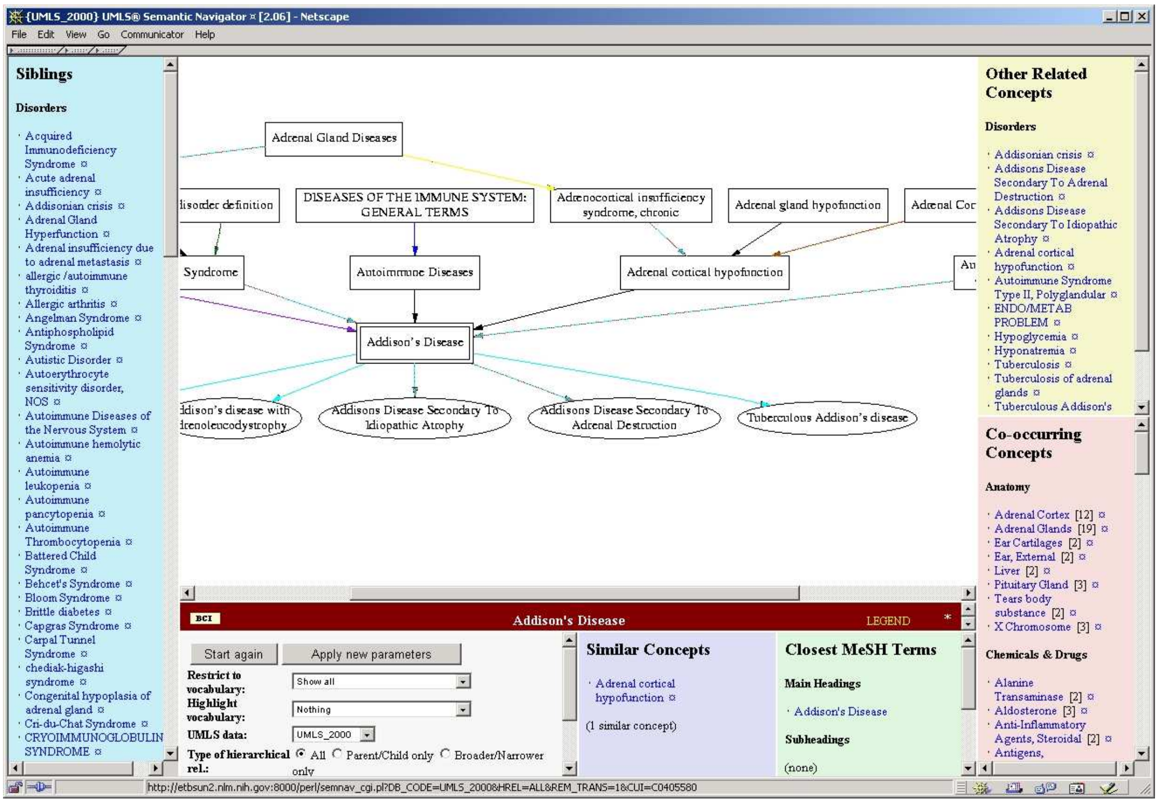


Figure 7 - UMLS Semantic Navigator

[4] www.research.att.com/sw/tools/graphviz/

[5] umlsks.nlm.nih.gov ► Resources ► Semantic Navigator

5.3 Research directions

As we said earlier, an important task of data visualization consists of selecting relevant data. This is especially true of co-occurring concepts. The distribution of the frequency of co-occurrence usually shows a long tail, which means that a large number of concepts co-occur a small number of times. As part of our study of co-occurrence of MeSH descriptors in MEDLINE [started with a focus on semantics, see 9], we plan to further analyze its statistical characteristics. An expected result is to restrict the list of co-occurring concepts to those that are relevant to the context.

We also plan to develop methods for visualizing the possible paths between concepts in the Metathesaurus, reflecting underlying polyhierarchical structures or, possibly, various combinations of hierarchical and associative relationships.

6 UTILIZE

Although many of the studies developed as part of the MEDICAL ONTOLOGY RESEARCH project have their own evaluation component, utilization of the methods and algorithms we created can be seen as an ultimate form of evaluation. Additionally, the utilization in other projects of the resources developed here allowed for rapid feedback and, in some cases, refinement.

6.1 Indexing Initiative

Our interest in semantic locality and semantic spaces was actually triggered by our participation to the Indexing Initiative project [29]. In this project, various methods are used to extract UMLS concepts are MEDLINE citations. These concepts are then restricted to MeSH descriptors which, once ranked, constitute a set of indexing terms for the document. Our task in this project was to restrict UMLS concepts to the MeSH vocabulary, using the principles of semantic locality on which the UMLS is based [30]. The major component of the Restrict to MeSH algorithm consists of building the graph of the ancestors of a given UMLS concept C, and selecting from this graph the concepts that belong to the MeSH vocabulary and whose semantic distance from C is minimal. This algorithm can be tuned from a strict mode (find only relevant MeSH terms, but with large number of failures) to a relaxed mode (find all possible MeSH terms, some of them being not relevant). The method used in the Indexing Initiative is a medium mode in which the selected MeSH terms are intended to be filtered and clustered according to additional information such as the frequency of each term in the source text and how often these terms co-occur in the medical literature. Related MeSH descriptors are found in 75% of all UMLS concepts. A table associating MeSH descriptors to UMLS concepts is produced each time a new version of the UMLS is released.

6.2 Clinical Trials Database

In patient- or consumer-oriented health information systems, such as the Clinical Trials database[6], condition terms are indexed by broad disease categories such as “Eye Diseases” or “Parasitic Diseases,” allowing users to navigate the system in browse mode. A condition term may be assigned to several disease categories, increasing the possibility of retrieving a given condition from different categories. For example, the term “adrenal medulla neoplasm” (a tumor of adrenal gland) could be assigned to both “Endocrine Diseases” and “Neoplasms” categories. Dynamic systems in which data may be added on a continuing basis require condition terms to be classified automatically, with no misclassified conditions and few non-classified conditions. We developed a method whereby specific disease names (or, more generally, names for medical conditions), referred to as condition terms, can be automatically classified into broad disease categories. Instead of using statistical classification techniques, we decided to exploit the semantic properties of the UMLS and to explore the possibility of using inter-concept relationships in the UMLS to select disease categories found in the semantic vicinity of a given condition term. This approach had already been used successfully in the Indexing Initiative (6.1).

In order to automatically classify condition terms into broad disease categories, we reused and combined three existing components: 1) Mapping terms to UMLS concepts; 2) Restricting UMLS concepts to MeSH; and 3) Mapping MeSH terms to disease categories. When applied to condition terms from the Clinical Trials database, this method assigned relevant categories to 92% of the 1823 condition terms encountered. 135 (7%) failed to be classified and 14 (.77%) were misclassified. This approach of using UMLS semantics for classification purposes was presented at the AMIA Annual Fall Symposium in 2000 [31].

6.3 Research directions

Several refinements could be explored to refine the Restrict to MeSH algorithm. Using the metaphor of stop-words in natural language processing, stop-concepts could be used for preventing wrong mappings from happening. Conversely, missing relationships in the Metathesaurus are known to prevent useful mappings from happening. For example, knowing that adjectives in the Metathesaurus are often represented as a distinct concept rather than clustered with their nominal equivalent, with no connection to it, we could use nominalization of adjectives for creating the missing links. Finally, although related MeSH descriptors based on the “other” relationships (essentially associative, not hierarchical) are less likely to be relevant, instead of not using them at all, we could take advantage of explicit mapping relationships recorded in the UMLS, one component of the “other” relationships.

Ultimately, when a semantic distance will be available in the UMLS, it should provide a quantification of semantic locality. The Restrict to MeSH algorithm is thus expected to take advantage of the semantic distance. We also plan to generalize this algorithm so that it finds the closest concepts not only in MeSH, but in any vocabulary integrated in the UMLS.

[6] *ClinicalTrials.gov*

Other projects developed at NLM could take advantage of an improved representation of the semantics of biomedical concepts. For example, in the Semantic Knowledge Representation project (SemRep), only relationships among semantic types in the Semantic Network are used for reasoning. Collaboration between SemRep and MEDICAL ONTOLOGY RESEARCH is planned to study how interpretation could take advantage of selected inter-concept relationships from the Metathesaurus as well. More generally, we expect to exploit the methods developed in this project for providing the background knowledge needed for semantic representation and knowledge processing.

7 Future plans

Among the multiple research directions mentioned earlier toward developing methods whereby ontologies could be acquired from existing resources, we set the following priorities:

- Semantic distance, as a quantification of semantic locality will be researched in priority since it has practical implications in other projects such as the Indexing Initiative. A model is expected within a year.
- Compositionality and other issues related to what types need to be represented in an ontology of the biomedical domain are mid-term goals.
- Finally, validation against other knowledge sources and evaluation of a better-organized knowledge structures for biomedical knowledge processing.

Tasks such as exploring other systems or investigating alternative methods for knowledge acquisition will be conducted in parallel, as needed.

8 Summary

The UMLS is an extensive source of biomedical concepts. It also provides a large number of inter-concept relationships and qualifies for a source of semantic spaces in the biomedical domain. However, the organization of knowledge in the UMLS is not principled nor consistent enough for it to qualify as an ontology of the biomedical domain. In the tradition of the UMLS, the approach we propose for going toward an ontology consists of refining the definition and organization of the existing semantic space. Both basic and applied research is needed to augment and better organize knowledge in the UMLS. A sound, ontological representation of biomedical knowledge is expected to enable tasks such as reasoning, currently hardly possible with the UMLS, while improving the performance of tasks already supported (e.g., information retrieval).

9 References

1. Sowa JF. *Knowledge representation : logical, philosophical, and computational foundations*. Pacific Grove, Ca.: Brooks/Cole; 2000.
2. Gruber TR. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition* 1993;5(2):199-220.
3. Chute CG. Clinical classification and terminology: some history and current observations. *J Am Med Inform Assoc* 2000;7(3):298-303.
4. Lindberg DA, Humphreys BL, McCray AT. The Unified Medical Language System. *Methods Inf Med* 1993;32(4):281-91.
5. UMLS. *UMLS Knowledge Sources*. 12th ed. Bethesda (MD): National Library of Medicine; 2001.
6. Landauer TK, Dumais ST. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 1997;104(2):211-240.
7. **Bodenreider O**, Bean CA. Relationships among knowledge structures: Vocabulary integration within a subject domain. In: Bean CA, Green R, editors. *Relationships in the organization of knowledge*: Kluwer; 2001. p. 81-98.
8. Burgun A, **Bodenreider O**. Aspects of the taxonomic relation in the biomedical domain. In: Welty C, editor. Second international conference on formal ontologies in information systems; 2001; Ogunquit, Maine; 2001. p. (submitted).
9. Burgun A, **Bodenreider O**. Methods for exploring the semantics of the relationships between co-occurring UMLS concepts. *Medinfo* 2001:(submitted).
10. Delugach HS. An exploration into semantic distance. In: Pfeiffer HD, Nagle TE, editors. 7th annual Workshop "Conceptual structures: theory and implementation"; 1992; Las Cruces, NM, USA, July 8-10, 1992: Springer-Verlag; 1992. p. 119-124.
11. Cooper MC. Semantic distance measures. *Computational Intelligence* 2000;16(1):79-93.
12. Rada R, Mili H, Bicknell E, Blettner M. Development and application of a metric on semantic nets. *IEEE Transactions on Systems, Man and Cybernetics* 1989;19(1):17-30.
13. Landauer TK, Foltz PW, Laham D. An introduction to latent semantic analysis. *Discourse Processes* 1998;25(2&3):259-284.
14. **Bodenreider O**, Burgun A, Botti G, Fieschi M, Le Beux P, Kohler F. Evaluation of the Unified Medical Language System as a medical knowledge source. *J Am Med Inform Assoc* 1998;5(1):76-87.
15. Cimino JJ. Auditing the Unified Medical Language System with semantic methods. *J Am Med Inform Assoc* 1998;5(1):41-51.
16. **Bodenreider O**. Circular Hierarchical Relationships in the UMLS: Etiology, Diagnosis, Treatment, Complications and Prevention. *Proc AMIA Symp* 2001:(submitted).

17. McCray AT, **Bodenreider O**. A conceptual framework for the biomedical domain. In: Sung M, Green R, editors. *Semantics of Relationships*: Kluwer; 2001. p. (to appear).
18. Burgun A, **Bodenreider O**. Comparing terms, concepts and semantic classes in WordNet and the Unified Medical Language System. *Proc NAACL Workshop, "WordNet and Other Lexical Resources: Applications, Extensions and Customizations"* 2001:(to appear).
19. Burgun A, **Bodenreider O**. Mapping the UMLS Semantic Network into General Ontologies. *Proc AMIA Symp* 2001:(to appear).
20. **Bodenreider O**, Burgun A, Rindflesch TC. Lexically-suggested hyponymic relations among medical terms and their representation in the UMLS. *Proc "Terminology and Artificial Intelligence"* 2001:(to appear).
21. McCray AT, **Bodenreider O**, Malley J, Browne AC. Evaluating UMLS Strings for Natural Language Processing. *Proc AMIA Symp* 2001:(submitted).
22. Sheretz DD, Tuttle MS, Cole WG, Erlbaum MS, Olson NE, Nelson SJ. A HyperCard implementation of Meta-1: the first version of the metathesaurus. *Proceedings of the 13th Annual Symposium on Computer Applications to Medical Care* 1989:1017-1018.
23. McCray AT, Razi AM, Bangalore AK, Browne AC, Stavri PZ. The UMLS Knowledge Source Server: a versatile Internet-based research tool. *Proc AMIA Annu Fall Symp* 1996:164-8.
24. McCray AT, Nelson SJ. The representation of meaning in the UMLS. *Methods Inf Med* 1995;34(1-2):193-201.
25. Nelson SJ, Tuttle MS, Cole WG, Sherertz DD, Sperzel WD, Erlbaum MS, et al. From meaning to term: semantic locality in the UMLS Metathesaurus. *Proc Annu Symp Comput Appl Med Care* 1991:209-13.
26. **Bodenreider O**. An object-oriented model for representing semantic locality in the UMLS. *Medinfo* 2001:(to appear).
27. McCray AT, Burgun A, **Bodenreider O**. Aggregating UMLS semantic types for reducing conceptual complexity. *Medinfo* 2001:(to appear).
28. **Bodenreider O**. A semantic navigation tool for the UMLS. *Proc AMIA Symp* 2000:971.
29. Aronson AR, **Bodenreider O**, Chang HF, Humphrey SM, Mork JG, Nelson SJ, et al. The NLM Indexing Initiative. *Proc AMIA Symp* 2000:17-21.
30. **Bodenreider O**, Nelson SJ, Hole WT, Chang HF. Beyond synonymy: exploiting the UMLS semantics in mapping vocabularies. *Proc AMIA Symp* 1998:815-9.
31. **Bodenreider O**. Using UMLS semantics for classification purposes. *Proc AMIA Symp* 2000:86-90.

10 Appendix: Team members

- Olivier Bodenreider, LHNCBC/CgSB
- Anita Burgun, LHNCBC/CgSB (Guest researcher, 2000-2001)

Olivier Bodenreider, M.D., Ph.D.
Staff Scientist

Education and Training

Institution	Degree	Years	Field of Study
University of Strasbourg, France, School of Medicine	MD	1980-1990	Medicine (including residency)
University of Nancy, France, School of Medicine	Research degree	1988-1991	Informatics, Statistics and Epidemiology
ISIAL, Henri Poincaré University, Nancy, France	Master degree	1989-1990	Computer Science
Henri Poincaré University, Nancy, France	Ph.D.	1991-1993	Medical Informatics
University of Nancy, France, School of Medicine	Master degree	1993-1994	Medical Information

Research and Professional Experience

Research

October 2000-present	**MEDICAL ONTOLOGY RESEARCH** Definition, Organization, visualization, and utilization of semantic spaces in the biomedical domain
October 1996-present	**INDEXING INITIATIVE** Mapping of UMLS concepts to MeSH, based on semantic locality
March 1994- October 1996	**MAOUSSC** Conceptual modeling for the description of medical procedures
July 1990-June 1993	**TOXICINE** Modeling of drug kinetics in clinical toxicology

Employment

October 2000-present	Staff Scientist, National Library of Medicine, Bethesda, MD
October 1997 September 2000	Senior System Analyst (MSD, Vienna, VA) Contractor at the National Library of Medicine, Bethesda, MD
October 1996- September 1997	Sabbatical year National Library of Medicine, Bethesda, MD
November 1991- October 1997	Assistant professor (medical informatics and biostatistics) University of Nancy, France, School of Medicine
	Attending Physician, Department of Medical Information University hospital (CHU), Nancy, France

Honors

Grants	3M (Health Care Division), Research Grant, 1996
	Association des Utilisateurs de Nomenclatures Nationales et Internationales en Santé, Research Grant, 1996
	Collège des Praticiens Spécialistes en Information et Communication Médicales, Research Grant, 1996
Professional membership	AMIA (American Medical Informatics Association)
Review Committees	AMIA, MEDINFO, Medical Informatics Europe

Selected Publications

1. **Bodenreider O**, Nelson SJ, Hole WT, Chang HF. Beyond synonymy: exploiting the UMLS semantics in mapping vocabularies. *Proc AMIA Symp* 1998:815-9.
2. **Bodenreider O**, McCray AT. From French vocabulary to the Unified Medical Language System: a preliminary study. *Medinfo* 1998;9(Pt 1):670-4.
3. Bouchet C, **Bodenreider O**, Kohler F. Integration of the analytical and alphabetical ICD10 in a coding help system. Proposal of a theoretical model for the ICD representation. *Medinfo* 1998;9(Pt 1):176-9.
4. Burgun A, **Bodenreider O**, Denier P, Delamarre D, Botti G, Oberlin P, et al. A collaborative approach to building a terminology for medical procedures using a Web-based application: from specifications to daily use. *Medinfo* 1998;9(Pt 1):596-9.
5. **Bodenreider O**, Burgun A, Botti G, Fieschi M, Le Beux P, Kohler F. Evaluation of the Unified Medical Language System as a medical knowledge source. *J Am Med Inform Assoc* 1998;5(1):76-87.
6. Aronson AR, **Bodenreider O**, Chang HF, Humphrey SM, Mork JG, Nelson SJ, et al. The NLM Indexing Initiative. *Proc AMIA Symp* 2000:17-21.
7. **Bodenreider O**, Zweigenbaum P. Identifying proper names in parallel medical terminologies. *Stud Health Technol Inform* 2000;77:443-7.
8. **Bodenreider O**. Using UMLS semantics for classification purposes. *Proc AMIA Symp* 2000:86-90.
9. **Bodenreider O**. A semantic navigation tool for the UMLS. *Proc AMIA Symp* 2000:971.
10. **Bodenreider O**. Comment les usagers accèdent à l'information médicale aux USA: l'exemple de MEDLINEplus [How consumers access health information in the USA: the MEDLINEplus example]. *Innovation et Technologie en Biologie et Médecine* 2000;21(5):286-90.
11. Jacquelinet C, **Bodenreider O**, Burgun A. Modelling Syllepse in Medical Knowledge Bases with application in the Domain of Organ Failure and Transplantation. In: OntoLex 2000 (Workshop on Ontologies and Lexical Knowledge Bases). Sozopol, Bulgaria, Sep 8-10; 2000.