A Lexical Metaschema for the UMLS Semantic Network

Li Zhang, Yehoshua Perl, Michael Halper[2], James Geller, George Hripcsak[3]

Computer Science Dept.
NJIT
Newark, NJ 07102
{lxz1853, yehoshua.perl, james.geller}@njit.edu

[2]Mathematics &
Computer Science Dept.
Kean University
Union, NJ 07083
mhalper@kean.edu

[3]Dept. of Medical Informatics
Columbia University
New York, NY 10032
hripcsak@columbia.edu

Reprint requests & all communications to:

Yehoshua Perl, Ph.D.
yehoshua.perl@njit.edu
CS Dept., NJIT
NJIT, University Heights, Newark, 07102
phone: (973) 596-3392
fax: (973) 642-7029

Abstract

Objective: A metaschema is a high-level abstraction network of the UMLS's Semantic Network (SN) obtained from a partition of the SN's collection of semantic types. Every metaschema has nodes, called meta-semantic types, each of which denotes a group of semantic types constituting a subject area of the SN. A new kind of metaschema, called the *lexical metaschema*, is derived from a lexical partition of the SN. The lexical metaschema is compared to previously derived metaschemas, e.g., the cohesive metaschema.

Design: A new lexical partitioning methodology is presented based on identical word-usage among the names of semantic types and the definitions of their respective children. The lexical metaschema is derived from the application of the methodology. We compare the constituent meta-semantic types and their underlying semantic-type groups with the previously derived cohesive metaschema. A similar comparison of the lexical partition and a published partition of the SN is also carried out.

Results: The lexical partition of the SN has 21 semantic-type groups, each of which represents a subject area. The lexical metaschema thus has 21 meta-semantic types, 19 meta-child-of hierarchical relationships, and 86 meta-relationships. Our comparison shows that 15 out of the 21 meta-semantic types in the lexical metaschema also appear in the cohesive metaschema, and 80 semantic types are covered by identical meta-semantic types or refinements between the two metaschemas. The comparison between the lexical partition and the semantic partition shows that they have very low similarity.

Conclusion: The algorithmically derived lexical metaschema serves as an abstraction of the SN and provides views representing different subject areas. It compares favorably with the cohesive metaschema derived via the SN's relationship configuration.

Keywords: Lexical Partition, Metaschema, String Matching, UMLS, Semantic Network

1 Introduction

The Unified Medical Language System (UMLS) [1, 2, 3] was designed by the National Library of Medicine (NLM) to overcome problems arising from discrepancies between terminologies used by various health care systems. It consists of concepts which reside in a repository called the Metathesaurus (META) [4, 5]. The Semantic Network (SN) part of the UMLS provides an overarching abstraction of the META [6]. The SN, consisting of 135 nodes called semantic types [7], is organized as a pair of trees rooted at **Event** and **Entity**,[1] respectively. The links of the two trees represent IS-A relationships, each of which connects a child semantic type to its parent. Besides the IS-A relationships, there are about 7,000 occurrences of 53 kinds of other non-IS-A semantic relationships (called, in short, semantic relationships) connecting pairs of semantic types.

While the SN is an important abstraction of the META, it is still a difficult source to peruse for orientation purposes due to its extensive content. To give an idea of the SN's complexity, its **Event** subnetwork is shown in Figure 1. Note that the figure displays neither the incoming relationships from semantic types out of the scope of the figure (i.e., from the **Entity** tree) nor the inherited relationships of the semantic types. Relationships to semantic types in the **Entity** hierarchy appear as named (or numbered) arrows without target semantic types. This figure clearly demonstrates the need to provide comprehensible access to the SN through simpler and more compact views to help user orientation. In previous work [8], we introduced the notion of metaschema, a higher-level

[1] A bold font will be used for semantic types, except in tables.

network derived from a partition of the SN [9]. Every metaschema serves as an abstraction of the SN. As shown in [8], a metaschema helps to generate various compact (partial) views that can help users in their orientation to the SN. Additional applications were described in [8, 10]. In [11], the notion of metaschema was extended from the SN to encompass a directed acyclic graph (DAG) semantic network. Two metaschemas were obtained for the DAG-structured Enriched Semantic Network (ESN) [12, 13].

In this paper, we introduce a new kind of lexical partitioning technique based on string matching from definitions of semantic types to the names of their parents. In this technique, a child and parent that are "lexically related" will be grouped together in the same group of the lexical partition. A metaschema, called the *lexical metaschema*, is then derived based on the lexical partition.

In this paper, we will compare the lexical metaschema to the previously derived *cohesive metaschema* [8]. This cohesive metaschema was derived based on the relationship structures of semantic types. For details of the cohesive metaschema, see Section 2.2. We also compare the lexical partition to the semantic partition presented in [14] because there is no metaschema for the semantic partition. For details of the semantic partition, see Section 2.3.

2 Background

2.1 A Metaschema of the SN

The notion of metaschema was introduced in [8] as an abstraction of the SN. A metaschema is based on a partition of the SN where the SN's IS-A hierarchy is partitioned into disjoint *semantic-type groups*. A semantic-type group is called *connected* if its semantic types together with their respective IS-A links constitute a connected subgraph of the SN with a unique root. A partition

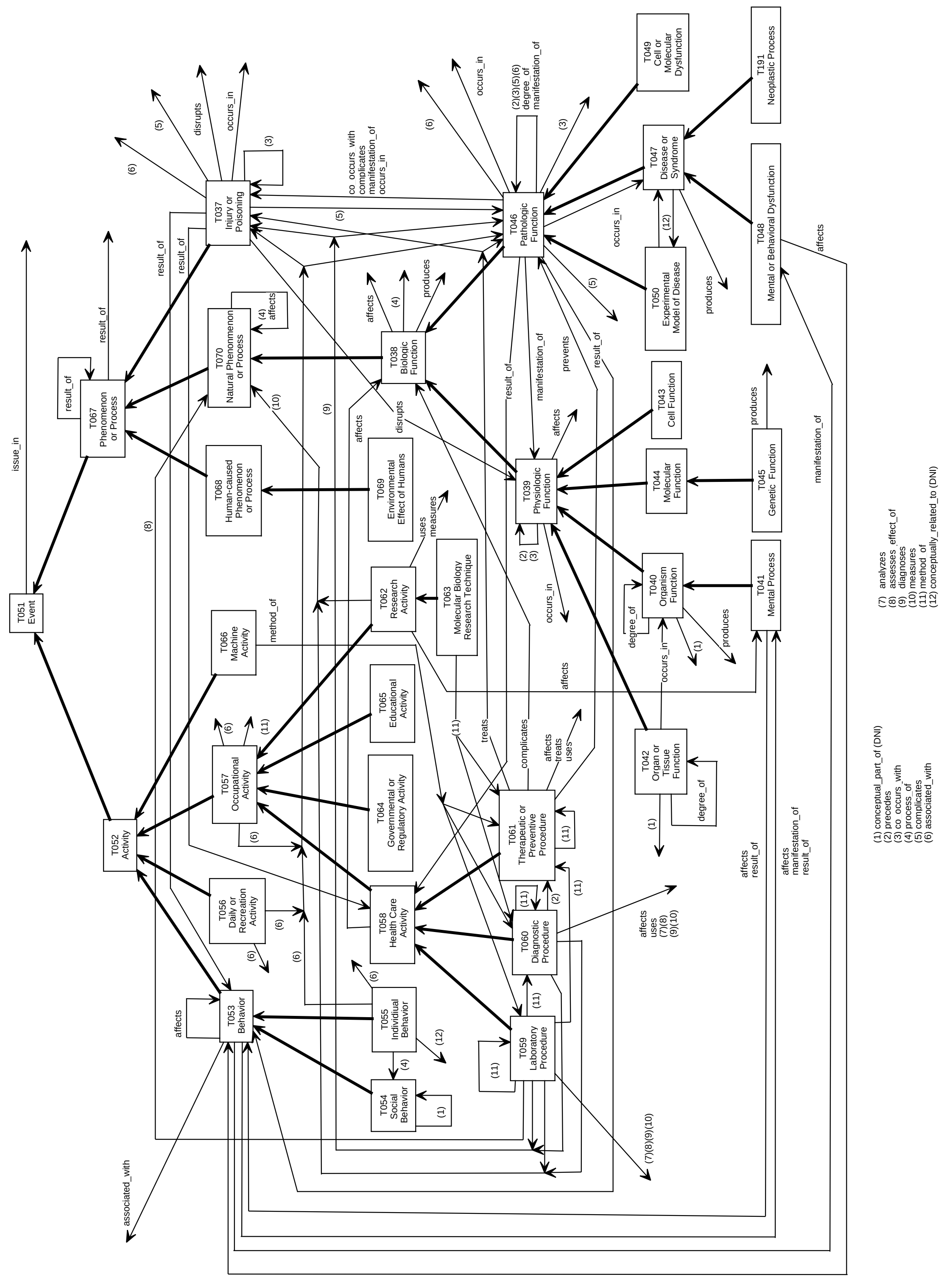


Figure 1: **Event** subnetwork of the SN

is called *connected* if each of its semantic-type groups is connected. A metaschema is based on a connected partition of the SN. Additionally, while a semantic-type group can be a singleton (i.e., or group of one semantic type), that semantic type cannot be a leaf in the SN's hierarchy. This condition is imposed because the metaschema should manifest some size reduction, which singletons do not contribute to. However, a singleton containing a non-leaf semantic type is allowed, since it may express an important internal branching point in the metaschema.

In a metaschema, each semantic-type group of the partition is represented by a single node, called a *meta-semantic type*. Two kinds of relationships connect meta-semantic types. The hierarchical *meta-child-of* relationships between meta-semantic types are derived as abstractions of the SN's IS-A links. The non-hierarchical relationships, called *meta-relationships*, are derived from the SN's semantic (non-hierarchical) relationships. The *meta-child-of* hierarchy supports inheritance of meta-relationships. Details of these derivations were presented in [8, 11], and a summary appears in Section 3.2.

For example, the hierarchy of the **Event** portion could be partitioned into the five semantic-type groups shown in Figure 2. Each semantic-type group is represented by a meta-semantic type in the corresponding metaschema. A meta-semantic type PHENOMENON OR PROCESS[2] is defined to represent the semantic-type group rooted at **Phenomenon or Process** in Figure 2. The metaschema hierarchy derived from the partition in Figure 2 is shown in Figure 3.

Overall, a diagram of a metaschema serves as a good visualization mechanism supporting orientation to the SN and, in turn, the META, and helps in the navigation of the UMLS knowledge. In [8] we introduced various partial graphical views of groups of semantic types supported by the

[2]Meta-semantic types will be written in "small caps" style.

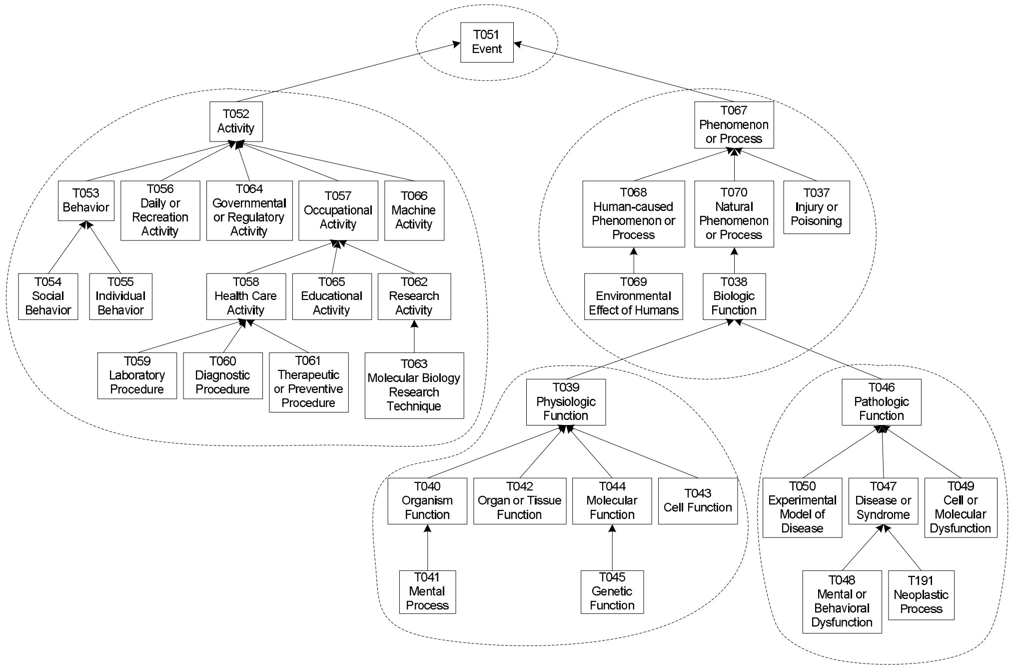


Figure 2: A partition example of the **Event** hierarchy of the SN

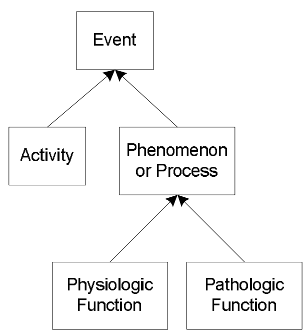


Figure 3: Metaschema hierarchy corresponding to the partition of the **Event** hierarchy of Figure 2

metaschema paradigm. These views can help in orientation of a user to the full scope of the SN's semantic relationships. In addition to the notion of metaschema, other previous work has focused on different methods to facilitate UMLS knowledge comprehension and visualization. Bodenreider and McCray described how to use visualization of semantic relationships as important indicators to explore coherence of semantic groups and help in auditing and validating the SN [15]. In [16], Nelson, *et al.*, presented the Hypercard browser MetaCard to enable users to extend the browsing process from META to a variety of different knowledge sources. In [17], knowledge exploration tools using levels of indentation to represent items standing in hierarchical relationships were used for displaying biomedical hierarchies in environments such as Protégé-2000. A review of knowledge visualization and navigation in the medical domain was presented by Tuttle *et al.* in [18].

2.2 The Cohesive Metaschema

We derived the cohesive metaschema in [8] based on the cohesive partition of the SN [9] into semantic-type groups satisfying semi-structural uniformity and semantical coherence. The cohesive partition, in fact, was an enhanced partition of the structural partition [9]. We defined the relationship structure of a semantic type to be the set of semantic relationships emanating from that semantic type. Therefore, in the structural partition, semantic types exhibiting the same relationship structure were grouped in the same semantic-type group. Hence, all semantic types in each semantic-type group in the structural partition are structurally uniform. A semantic-type group which has only one semantic type is called a singleton. The resulting structural partition contained 71 semantic-type groups. Among the 71 structural groups, 47 are singletons.

An effective partition should reflect not only the structure of semantic types, but also the semantics of those types. One way to guarantee the semantic coherence of a group is to make sure all the semantic types in the group are subsumed under one category, that is, the group has one unique root. Some semantic-type groups in the structural partition had two roots, which means that these groups were not semantically coherent. Furthermore, the large number of singletons made the structural partition not proper for defining a metaschema.

To address these two problems of the structural partition, rules were defined in [9] to merge leaf singletons into the groups of their respective parents and to enforce that each semantic-type group have a unique root semantic type. In this way an enhanced partition containing 28 semantic-type groups was obtained, referred to as the cohesive partition.

Based on the cohesive partition, the cohesive metaschema was derived, containing 28 meta-semantic types, 26 *meta-child-of* relationships, and 133 meta-relationships. Its hierarchy consists of two trees rooted at the ENTITY and EVENT meta-semantic types, similar to the situation in the IS-A hierarchy of the SN. However these two trees are more compact than the corresponding trees of the SN.

2.3 The Semantic Partition

In [14], McCray, Burgun, and Bodenreider presented a partition of the SN into 15 groups, with each group representing a subject area. This partition was derived externally since the authors first picked different subject areas in medicine and then assigned each semantic type to a proper subject area. The groupings of semantic types were subject to a set of general principles including, *semantic validity* (the groups must be semantically coherent); *parsimony* (the number of groups should be as small as possible); *exclusivity* (each semantic type must belong to only one group); *completeness* (the groups must cover the full domain); *naturalness* (the groups characterize the domain in a way that is acceptable to a domain expert); and *utility* (the groups must be useful for some purpose). Table 1 shows the 15 groups resulting from applying these rules. Two possible methods were presented to measure the degree of semantic coherence for each group in the resulting partition. One way is too see if all semantic types in a group are hierarchically related to each other. The other way is to analyze the semantic relationships exhibited by semantic types in a given group. The resulting partition can be used for display purposes to reduce conceptual complexity and to provide a broad overview of the SN. It might be also helpful to discover inconsistencies in the representation of the SN.

3 Methods

We first introduce our lexical partitioning technique for generating a lexical partition. Then we describe how to derive the lexical metaschema based on the lexical partition. We further present the comparison techniques used to compare the lexical metaschema to the cohesive metaschema in [8]. Similar techniques are used to compare the lexical partition and the semantic partition in [14].

3.1 A Lexical Partitioning Technique Based on String Matching

Our lexical partitioning technique is based on string matches among pairs of child and parent semantic types. We define the notion of "string match" in the following, where we use the term "child/parent pair" ("CP-pair" for short) to denote a pair of semantic types ($\mathbf{T}_1$, $\mathbf{T}_2$) such that $\mathbf{T}_1$ is a child of $\mathbf{T}_2$ in the SN.

Definition (String Match): Let ($\mathbf{T}_1$, $\mathbf{T}_2$) be a CP-pair. A string match between $\mathbf{T}_1$ and $\mathbf{T}_2$ is a triple ($\mathbf{T}_1$; $\mathbf{T}_2$; S) such that S is a string appearing both in the definition of $\mathbf{T}_1$ and the name of $\mathbf{T}_2$. S is called the common string and must contain one or more (not necessarily consecutive) complete words (ignoring case). □

For example, the definition of **Plant** contains the word "organism" which happens to be the name of its parent **Organism**. Hence, there is a string match (**Plant**; **Organism**; "organism"). On the other hand, there is no string match from **Biologically Active Substance** to its parent **Chemical Viewed Functionally**.

From this overlapping word usage, we define:

Table 1: Partition of the SN Presented in [14]

Group	Semantic Types in Group
Activities and Behaviors	Occupational Activity; Behavior; Activity; Event; Individual Behavior; Daily or Recreational Activity; Governmental or Regulatory Activity; Machine Activity; Social Behavior
Anatomy	Body Location or Region; Body System; Body Part, Organ, or Organ Component; Anatomical Structure; Embryonic Structure; Tissue; Cell; Body Space or Junction; Fully Formed Anatomical Structure; Body Substance; Cell Component
Chemicals and Drugs	Biomedical or Dental Material; Biologically Active Substance; Organic Chemical; Pharmacologic Substance; Chemical; Enzyme; Neuroreactive Substance or Biogenic Amine; Hormone; Chemical Viewed Structurally; Vitamin; Immunologic Factor; Indicator, Reagent, or Diagnostic Aid; Clinical Drug; Inorganic Chemical; Element, Ion, or Isotope; Antibiotic; Hazardous or Poisonous Substance; Receptor; Steroid; Eicosanoid; Chemical Viewed Functionally; Nucleic Acid, Nucleoside, or Nucleotide; Organophosphorus Compound; Amino Acid, Peptide, or Protein; Carbohydrate; Lipid
Concepts and Ideas	Classification; Quantitative Concept; Qualitative Concept; Temporal Concept; Idea or Concept; Conceptual Entity; Group Attribute; Language; Intellectual Product; Spatial Concept; Functional Concept; Regulation or Law
Devices	Research Device; Medical Device
Disorders	Finding; Injury or Poisoning; Pathologic Function; Experimental Model of Disease; Disease or Syndrome; Sign or Symptom; Anatomical Abnormality; Neoplastic Process; Mental or Behavioral Dysfunction; Cell or Molecular Dysfunction; Acquired Abnormality; Congenital Abnormality
Genes and Molecular Sequences	Molecular Sequence; Amino Acid Sequence; Carbohydrate Sequence; Nucleotide Sequence; Gene or Genome
Geographic Areas	Geographic Area
Living Beings	Organism; Population Group; Fungus; Alga; Virus; Human; Plant; Archaeon; Group; Professional or Occupational Group; Reptile; Family Group; Age Group; Patient or Disabled Group; Rickettsia or Chlamydia; Amphibian; Mammal; Fish; Bird; Animal; Vertebrate; Invertebrate; Bacterium
Objects	Entity; Manufactured Object; Physical Object; Substance; Food
Occupations	Occupation or Discipline; Biomedical Occupation or Discipline
Organizations	Health Care Related Organization; Self-help or Relief Organization; Professional Society; Organization
Phenomena	Phenomenon or Process; Human-caused Phenomenon or Process; Laboratory or Test Result; Natural Phenomenon or Process; Biologic Function; Environmental Effect of Humans
Physiology	Organism Attribute; Cell Function; Organ or Tissue Function; Organism Function; Genetic Function; Molecular Function; Physiologic Function; Mental Process; Clinical Attribute
Procedures	Laboratory Procedure; Health Care Activity; Molecular Biology Research Technique; Diagnostic Procedure; Educational Activity; Research Activity; Therapeutic or Preventive Procedure

Definition (Lexically related): A CP-pair ($\mathbf{T}_1$, $\mathbf{T}_2$) is said to be lexically related if there exists a string match between $\mathbf{T}_1$ and $\mathbf{T}_2$. □

For example, the CP-pair (**Plant**, **Organism**) is lexically related, while (**Biologically Active Substance**, **Chemical Viewed Functionally**) is not lexically related. The child in a CP-pair that is not lexically related is called *lexically independent* from its parent. The two roots of the SN, **Entity** and **Event**, are also called lexically independent since they do not have parents to which they may be related.

In order for a metaschema to help users in their orientation to the SN, its associated partition must have semantic-type groups that capture various subject areas within the medical field. An underlying assumption of our lexical partitioning technique is that if a CP-pair is lexically related, then both semantic types belong to the same subject area and should therefore be in the same semantic-type group. If, on the other hand, a CP-pair is not lexically related, then the child can be seen as a transition to a new (although still hierarchically related) subject area. The child in this case will be made the root of a new semantic-type group in the lexical partition. Its own lexically related children and, in turn, all their lexically related children, etc., will be part of this semantic-type group, too. The justification for the above assumption is that SN's semantic types' definitions are (expected to be) Aristotelian [19] definitions in which reference is made to the genus (here, the parent semantic type) and differences (differentiae) between child and parent. From a linguistic perspective, the sentence of the definition of a child semantic type often contains the name of its parent as its head. We are indebted to an anonymous reviewer for pointing this out. We interpret the occurrence of a string match to indicate a strong semantic connection of the child to the parent semantic type.

For example, **Biologically Active Substance**, which is lexically independent from its parent, will start a new subject area and thus will be the root of a new semantic-type group. In contrast, **Plant** is deemed to belong to the same subject area as **Organism**, and thus resides in the same semantic-type group.

To construct the lexical partition, we need to identify all lexically related CP-pairs. That is, we need to check if string matches exist for the 133 CP-pairs in the SN. In the following, we describe the partitioning process as a series of four steps.

Step 1: Preprocess the definitions and names of all CP-pairs. The preprocessing includes stop-word removal, verb-variant processing, and lexical normalization [13, 20].

Step 2: Apply the "AllMatches" algorithm (originally defined in [13]) to the preprocessed CP-pairs to identify all string matches in the SN.

Step 3: For each CP-pair, if there exists a string match, mark the CP-pair as "lexically related"; otherwise, mark it as "lexically unrelated."

Step 4: For each lexically unrelated CP-pair: if the child is not a leaf, then the child marks the root of a new semantic-type group in the partition; otherwise, the leaf is assigned to the same semantic-type group as its parent. For each lexically related CP-pair: the child is assigned to the same semantic-type group as its parent.

We now review the AllMatches algorithm which is used in Step 2 to find string matches for all CP-pairs in the SN. The input file to the algorithm contains the names and definitions of semantic types after the preprocessing step.

In the description of the AllMatches algorithm, we assume that $\mathbf{T}_1$, $\mathbf{T}_2, \ldots, \mathbf{T}_m$ are all semantic types in the SN. (In the 2003 version, $m = 135$). We use the notation *DEF*($\mathbf{T}_i$) to represent the definition of the semantic type $\mathbf{T}_i$, $1 \le i \le 135$, after preprocessing. *NAME*($\mathbf{T}_i$) is used to represent the name of $\mathbf{T}_i$, in the form of a string, after preprocessing. For example, suppose $\mathbf{T}_i =$ **Anatomical Structure**, which is defined as: "a normal or pathological part of the anatomy or structural organization of an organism." After preprocessing, *NAME*($\mathbf{T}_i$) ="anatomy structure" and *DEF*($\mathbf{T}_i$) = "normal pathology part anatomy structure organization organism."

In the following AllMatches algorithm, we use a list L to hold all common strings. We also use the following functions defined for lists:

Length(): Return the number of elements in the list

Retrieve(k): Retrieve the k^{th} element of the list

AllMatches algorithm: Find all string matches in the SN.

```
For (i = 1 to m)
    For all T_j (1 < j < m & j ≠ i)
        If (T_j is a child of T_i)
        {   L = FindCommonStrings(DEF(T_j),NAME(T_i));
            //write string matches to the output file
            For (k = 1 to L.Length())
            {   S = L.Retrieve(k);   // get the k^th element of L
                write (T_j; T_i; S) to output file;
            }
        }
```

The function FindCommonStrings(R_1, R_2) is used to find all common strings involving a given pair of strings R_1 and R_2. During a call, R_1 is the definition of a semantic type $\mathbf{T}_j$ in a string format, and R_2 is the name of a semantic type $\mathbf{T}_i$ as a string. For each CP-pair ($\mathbf{T}_j$, $\mathbf{T}_i$) we call FindCommonStrings(*DEF*($\mathbf{T}_j$), *NAME*($\mathbf{T}_i$)) to get all possible common strings between *DEF*($\mathbf{T}_j$) and *NAME*($\mathbf{T}_i$). Each such common string is inserted into L. We say that a match is *redundant* if its constituent common string S is a substring of another match's common string (again ignoring case). FindCommonStrings(*DEF*($\mathbf{T}_j$), *NAME*($\mathbf{T}_i$)) identifies the redundant matches and does not return them. So, L contains no redundant common strings. Finally, all string matches ($\mathbf{T}_j$; $\mathbf{T}_i$; S) are written to the output file. After AllMatches has been executed, we have a file containing all string matches for all CP-pairs in the SN.

We must note that even though "lexically related" is not transitive, the following is a consequence of our rules. If (**B**, **A**) is a lexically related CP-pair, then **B** is assigned to the same semantic-type group as **A**. Meanwhile, if (**C**, **B**) is a lexically related CP-pair, then **C** is assigned to the same semantic-type group as **B**. Therefore, **A**, **B**, and **C** will be in the same semantic-type group in the lexical partition.

3.2 Metaschema Derivation

With the lexical partition in place, we can derive the lexical metaschema. A metaschema comprises three kinds of components: meta-semantic types, *meta-child-of* relationships, and meta-relationships. These are defined below along with their derivations.

A *meta-semantic type* is a node defined to represent a single semantic-type group. It is given the name of the semantic-type group's root. The root of the semantic-type group is, by definition, also called the *root of the meta-semantic type*. The size of a meta-semantic type is the number of semantic types in the group it represents.

A *meta-child-of* relationship ("*meta-child-of*" for short) is a link between two meta-semantic types representing an IS-A relationship between two semantic types of the corresponding semantic-type groups. More specifically, let **A** and $\mathbf{B}_r$ be semantic types in the semantic-type groups of meta-semantic types A and B, respectively. Furthermore, let $\mathbf{B}_r$ be the root of B and $\mathbf{B}_r$ IS-A **A**. Then in the metaschema, we define a *meta-child-of* directed from B to A. Note that the semantic type **A** does not need to be the root of its meta-semantic type. Only the source $\mathbf{B}_r$ has to be a root in order for a new *meta-child-of* to be induced in the metaschema. The derivation of the *meta-child-of* links is motivated in detail in [8].

A *meta-relationship* is a link between two meta-semantic types representing a specific semantic relationship (non-IS-A relationship) between the two corresponding semantic-type groups. Specifically, let $\mathbf{A}_r$ and **B** be semantic types in the semantic-type groups of meta-semantic types A and B, respectively. Furthermore, let $\mathbf{A}_r$ be the root of A and let there exist a semantic relationship *rel* from $\mathbf{A}_r$ to **B**. Then in the metaschema, there exists a meta-relationship `rel`[3] directed from A to B. Note that the semantic type **B** does not need to be the root of its meta-semantic type. Only the source of the relationship *rel* (i.e., $\mathbf{A}_r$) has to be a root in order for a new meta-relationship *rel* to be induced in the metaschema. We will now justify this choice.

Suppose there is a meta-relationship `rel` from a meta-semantic type A to a meta-semantic type B in the metaschema. Our interpretation of this meta-relationship is that for every semantic type $\mathbf{A}_i$ in the group represented by A, there must exist a semantic type $\mathbf{B}_j$ in the group represented by B, such that in the SN there is a semantic relationship *rel* from $\mathbf{A}_i$ to $\mathbf{B}_j$. This interpretation is proper only if in the SN *rel* exists from the root $\mathbf{A}_r$ of the meta-semantic type A, since each semantic type $\mathbf{A}_i$ in the group represented by A inherits this relationship (see Figure 4). However, this will not be true if in the SN *rela* is introduced at a non-root semantic type in the group of the meta-semantic type A, since such a relationship is not inherited to all semantic types $\mathbf{A}_i$ of the group represented by A, just to the descendants of $\mathbf{A}_i$. Thus `rela` should not occur in the metaschema.

[3]Meta-relationships will be written in a courier font.

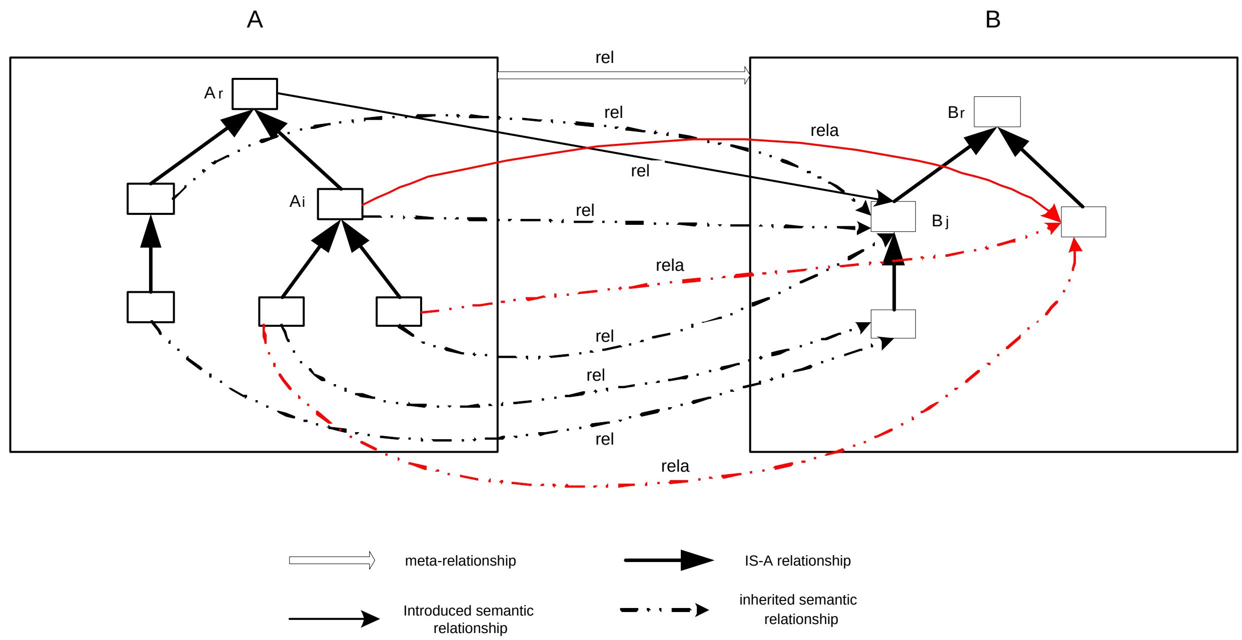


Figure 4: Interpretation of the definition of meta-relationship

3.3 Comparison Techniques

3.3.1 Comparison of Metaschemas

In our comparison of two metaschemas, we will consider not only the meta-semantic types' names but also the underlying semantic-type groups represented by the meta-semantic types. To support the comparison, we need some definitions.

Let M_1 and M_2 be two metaschemas of the SN.

Definition (Identical): A meta-semantic type A in M_1 is *identical* to a meta-semantic type B in M_2 if both meta-semantic types have the same underlying semantic-type group. □

Definition (Similar): A meta-semantic type A in M_1 is *similar* to a meta-semantic type B in M_2 if the roots of their underlying semantic-type groups are the same, while the groups are different. □

This definition implies that the names of two similar meta-semantic types are equal. To better understand the differences between pairs of similar meta-semantic types, we note that in some cases, the difference reflects various levels of granularity in the partition, rather than major disagreements between the metaschemas. A meta-semantic type in one metaschema may be split into several separate meta-semantic types in the other metaschema.

To be formal, we define "refinement" for meta-semantic types as follows. Let $G_M(\mathrm{A})$ denote the semantic-type group represented by the meta-semantic type A in the metaschema M.

Definition (Refinement): Let A be a meta-semantic type in metaschema M_1. If there exists a set of meta-semantic types $\{\mathrm{B}_1, \mathrm{B}_2, \ldots, \mathrm{B}_k\}$ ($k \geq 2$) in metaschema M_2 such that $G_{M_1}(\mathrm{A}) = \cup_{i=1}^{k} G_{M_2}(\mathrm{B}_i)$, then the set $\{\mathrm{B}_1, \mathrm{B}_2, \ldots, \mathrm{B}_k\}$ is called a *refinement* of A. □

In our comparison we will measure the percentage of identical and similar meta-semantic types and of their refinements.

3.3.2 Comparison of Partitions

The semantic partition in [14] is not a connected partition since some semantic-type groups have isolated semantic types or more than one tree. Therefore, there is no metaschema that can be derived from that semantic partition. Hence, we can only compare the lexical partition with the semantic partition. Furthermore, since the groups in the semantic partition are not named after semantic types, we cannot compare names of groups. We can only compare the semantic-type groups of the two partitions. To support the comparison, we need some definitions.

Let P_1 and P_2 be two partitions of the SN.

Definition (Identical): A semantic-type group A in P_1 is *identical* to a semantic-type group B in

P_2 if both semantic-type groups have the same set of semantic types. □

We only define similarity between two semantic-type groups that are connected. As mentioned before, we require that each connected group of the SN has a unique root.

Definition (Similar): A semantic-type group A in P_1 is *similar* to a semantic-type group B in P_2 if the roots of the two groups are equal and the set of semantic types in A is a subset of the set of semantic types in B or the set of semantic types in B is a subset of the set of semantic types in A. □

To better understand the differences between pairs of semantic-type groups, we note that in some cases the difference reflects various levels of granularity in the partition rather than major difference. A semantic-type group in one partition may be split into several separate semantic-type groups in the other partition.

We define "refinement" for semantic-type groups as follows. Let $S_P(\mathrm{A})$ denote the set of semantic types of the semantic-type group A in the partition P.

Definition (Refinement): Let A be a semantic-type group in partition P_1. If there exists a set of semantic-type groups $\{\mathrm{B}_1, \mathrm{B}_2, \ldots, \mathrm{B}_k\}$ ($k \geq 2$) in partition P_2 such that $S_{P_1}(\mathrm{A}) = \cup_{i=1}^{k} S_{P_2}(\mathrm{B}_i)$, then the set $\{\mathrm{B}_1, \mathrm{B}_2, \ldots, \mathrm{B}_k\}$ is called a *refinement* of the semantic-type group A. □

In our comparison we will measure percentage of identical and similar groups and of their refinements.

4 Results

4.1 Lexical Metaschema

Applying the "AllMatches" algorithm (Section 3.1) to the 133 CP-pairs results in string matches involving 88 CP-pairs. The string match (**Plant**; **Organism**; "organism") is one of them. Hence, about 70% of the children in the SN refer to the name or part of the name of their respective parents in their definitions. Therefore, there are 88 lexically related CP-pairs and 45 that are not lexically related.

In total, there are 47 lexically independent semantic types (including **Entity** and **Event**), among which 21 are non-leaf semantic types, and 26 are leaves. For example, the pair (**Organism**, **Physical Object**) is not lexically related, and **Organism** is a non-leaf semantic type. The pair (**Human**, **Mammal**) is not lexically related either, but **Human** is a leaf. Table 2 displays all 47 lexically independent semantic types. Each of the 21 non-leaf, lexically independent semantic types starts a new semantic-type group. Each of the 26 lexically independent leaves is assigned to the group of its respective parent. The two semantic types of each of the 88 lexically related CP-pairs are assigned to the same respective groups.

For example, **Organism** is a non-leaf, lexically independent semantic type; its child **Archaeon** is a lexically independent leaf; and the CP-pair (**Organism**, **Plant**) is lexically related. Hence, **Organism** starts a new semantic-type group; **Archaeon** and **Plant** are also assigned to this group. The chart in Figure 5 shows the distribution of the numbers of semantic-type groups according to their sizes. For example, there are four semantic-type groups of size six.

In Table 3 each row shows a root of a semantic-type group, the group's size, and the complete

Table 2: 47 lexically independent semantic types

Non-leaves	Leaves
T001 Organism T017 Anatomical Structure T032 Organism Attribute T033 Finding T039 Physiologic Function T046 Pathologic Function T051 Event T052 Activity T067 Phenomenon or Process T071 Entity T072 Physical Object T078 Idea or Concept T082 Spatial Concept T085 Molecular Sequence T090 Occupation or Discipline T092 Organization T109 Organic Chemical T119 Lipid T121 Pharmacologic Substance T123 Biologically Active Substance T167 Substance	T016 Human T022 Body System T024 Tissue T026 Cell Component T034 Laboratory or Test Result T037 Injury or Poisoning T048 Mental or Behavioral Dysfunction T050 Experimental Model of Disease T059 Laboratory Procedure T060 Diagnostic Procedure T061 Therapeutic or Preventive Procedure T083 Geographic Area T110 Steroid T111 Eicosanoid T114 Nucleic Acid, Nucleoside, or Nucleotide T116 Amino Acid, Peptide, or Protein T118 Carbohydrate T125 Hormone T126 Enzyme T131 Hazardous or Poisonous Substance T171 Language T184 Sign or Symptom T185 Classification T191 Neoplastic Process T192 Receptor T194 Archaeon

list of the semantic types in the group. For example, the semantic-type group rooted at **Organism** has 17 semantic types which are listed in the first row of the table. The groups are listed according to the order of their roots in Table 2.

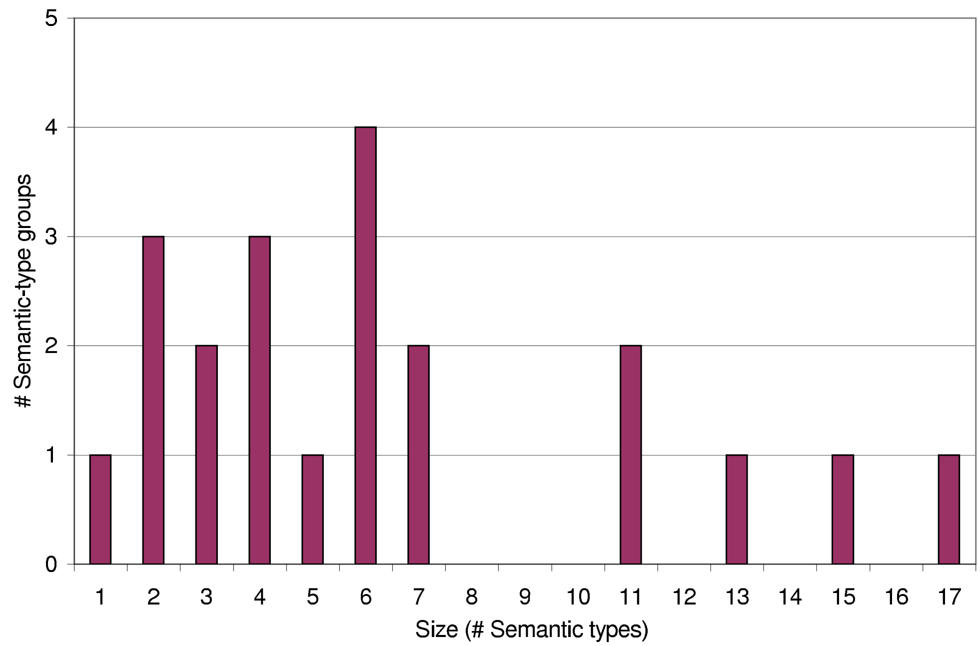


Figure 5: Size distribution of semantic-type groups

Root of Semantic-Type Group	Size	Semantic Types in Group
Organism	17	Organism; Plant; Alga; Archaeon; Virus; Animal; Invertebrate; Vertebrate; Mammal; Human; Reptile; Fish; Bird; Amphibian; Bacterium; Fungus; Rickettsia or Chlamydia
Anatomical Structure	11	Anatomical Structure; Embryonic Structure; Fully Formed Anatomical Structure; Body Part, Organ, or Organ Component; Tissue; Cell; Cell Component; Anatomical Abnormality; Acquired Abnormality; Congenital Abnormality; Gene or Genome
Organism Attribute	2	Organism Attribute; Clinical Attribute
Finding	3	Finding; Sign or Symptom; Laboratory or Test Result
Physiologic Function	7	Physiologic Function; Organ or Tissue Function; Organism Function; Mental Process; Molecular Function; Genetic Function; Cell Function
Pathologic Function	6	Pathologic Function; Experimental Model of Disease; Disease or Syndrome; Neoplastic Process; Mental or Behavioral Dysfunction; Cell or Molecular Dysfunction
Event	1	Event

Root of Semantic-Type Group	Size	Semantic Types in Group
Activity	15	Activity; Behavior; Individual Behavior; Social Behavior; Daily or Recreational Activity; Machine Activity; Occupational Activity; Health Care Activity; Laboratory Procedure; Diagnostic Procedure; Therapeutic or Preventive Procedure; Governmental or Regulatory Activity; Educational Activity; Research Activity; Molecular Biology Research Technique
Phenomenon or Process	6	Phenomenon or Process; Human-caused Phenomenon or Process; Environmental Effect of Humans; Natural Phenomenon or Process; Biologic Function; Injury or Poisoning;
Entity	13	Entity; Conceptual Entity; Group Attribute; Language; Intellectual Product; Classification; Regulation or Law; Group; Professional or Occupation Group; Population Group; Family Group; Age Group; Patient or Disabled Group
Physical Object	6	Physical Object; Manufactured Object; Research Device; Medical Device; Medical Delivery Device; Clinical Drug;
Idea or Concept	6	Idea or Concept; Functional Concept; Body System; Temporal Concept; Qualitative Concept; Quantitative Concept
Spatial Concept	4	Spatial Concept; Geographic Area; Body Location or Region; Body Space or Junction
Molecular Sequence	4	Molecular Sequence; Amino Acid Sequence; Carbohydrate Sequence; Nucleotide Sequence
Occupation or Discipline	2	Occupation or Discipline; Biomedical Occupation or Discipline
Organization	4	Organization; Professional Society; Health Care Related Organization; Self-help or Relief Organization
Lipid	3	Lipid; Steroid; Eicosanoid
Pharmacologic Substance	2	Pharmacologic Substance; Antibiotic
Biologically Active Substance	7	Biologically Active Substance; Receptor; Vitamin; Enzyme; Hormone; Neuroreactive Substance or Biogenic Amine; Immunologic Factor

Root of Semantic-Type Group	Size	Semantic Types in Group
Substance	11	Substance; Body Substance; Food; Chemical; Chemical Viewed Functionally; Hazardous or Poisonous Substance; Biomedical or Dental Material; Indicator, Reagent, or Diagnostic Aid; Chemical Viewed Structurally; Inorganic Chemical; Element, Ion, or Isotope
Organic Chemical	5	Organic Chemical; Amino Acid, Peptide, or Protein; Organophosphorus Compound; Nucleic Acid, Nucleoside, or Nucleotide; Carbohydrate
Total: 21	135	

Table 3: Lexical partition of the SN

With the lexical partition in place, the lexical metaschema can be derived. For example, a meta-semantic type PATHOLOGIC FUNCTION is defined to represent the semantic-type group rooted at **Pathologic Function**. PATHOLOGIC FUNCTION has six constituent semantic types. **Pathologic Function** is the root of PATHOLOGIC FUNCTION, and **Biologic Function** is in the group represented by PHENOMENON OR PROCESS. Since **Pathologic Function** IS-A **Biologic Function** in the SN, there exists a *meta-child-of* link directed from PATHOLOGIC FUNCTION to PHENOMENON OR PROCESS. Meanwhile, the four semantic relationships, `co_occurs_with`, `complicates`, `manifestation_of`, and `occurs_in`, are defined from **Pathologic Function**, the root of PATHOLOGIC FUNCTION, to **Injury or Poisoning**, which is in PHENOMENON OR PROCESS. Therefore, four meta-relationships, `co_occurs_with`, `complicates`, `manifestation_of`, and `occurs_in`, are defined from PATHOLOGIC FUNCTION to PHENOMENON OR PROCESS.

The hierarchy of the lexical metaschema is shown in Figure 6. The size of a meta-semantic type is displayed in parentheses following its name. Figure 7 shows the metaschema including all *meta-child-of*'s and meta-relationships. Overall, the metaschema contains 21 meta-semantic types, 19 *meta-child-of*'s, and 86 meta-relationships. The average size of a meta-semantic type is close to six.

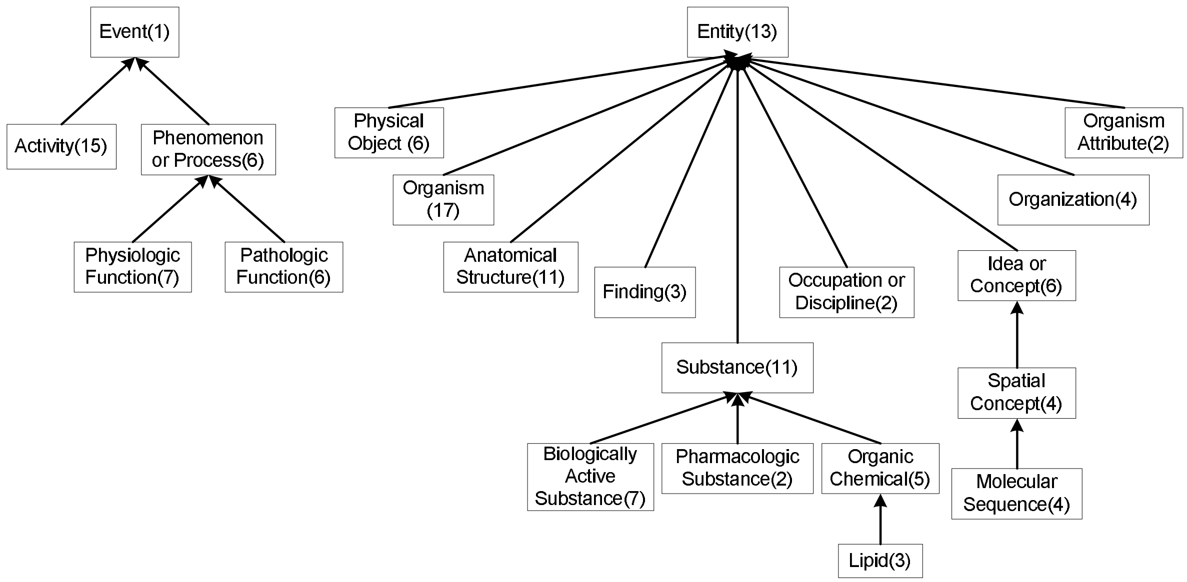


Figure 6: Lexical metaschema hierarchy

4.2 Comparison between the Cohesive Metaschema and the Lexical Metaschema

To facilitate the comparison between the cohesive and lexical metaschemas, we draw both their hierarchies in Figure 8. In both metaschemas, identical meta-semantic types are indicated by black shadows. Similar meta-semantic types are denoted by gray shadows. The number of semantic types in the group represented by an MST appears in the parenthesis following the MST name.

The cohesive metaschema contains 28 meta-semantic types, while the lexical metaschema contains 21 meta-semantic types. There are eight identical meta-semantic types between the two metaschemas. For example, PHYSIOLOGIC FUNCTION is a meta-semantic type in both metaschemas representing the same underlying semantic-type group containing seven semantic

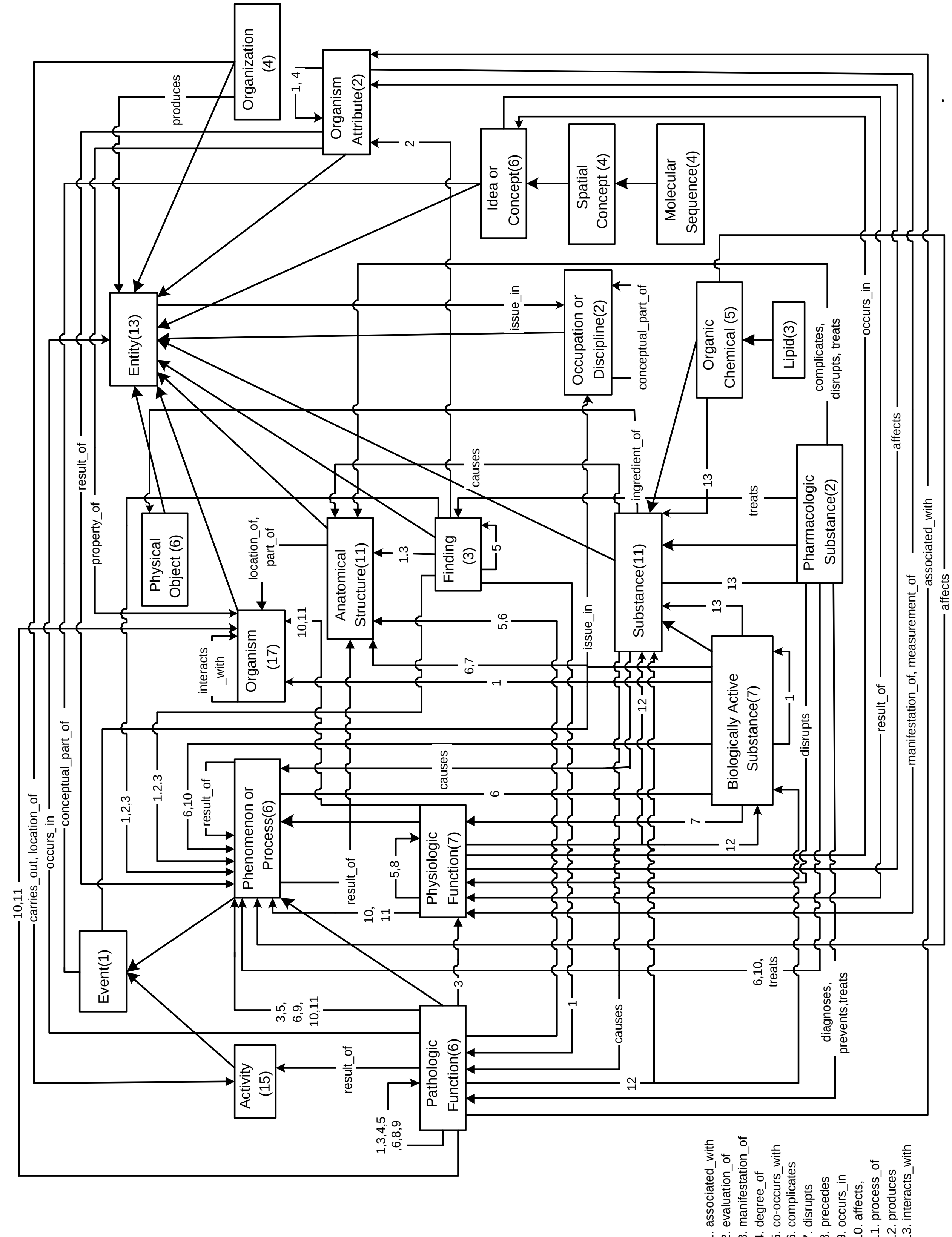
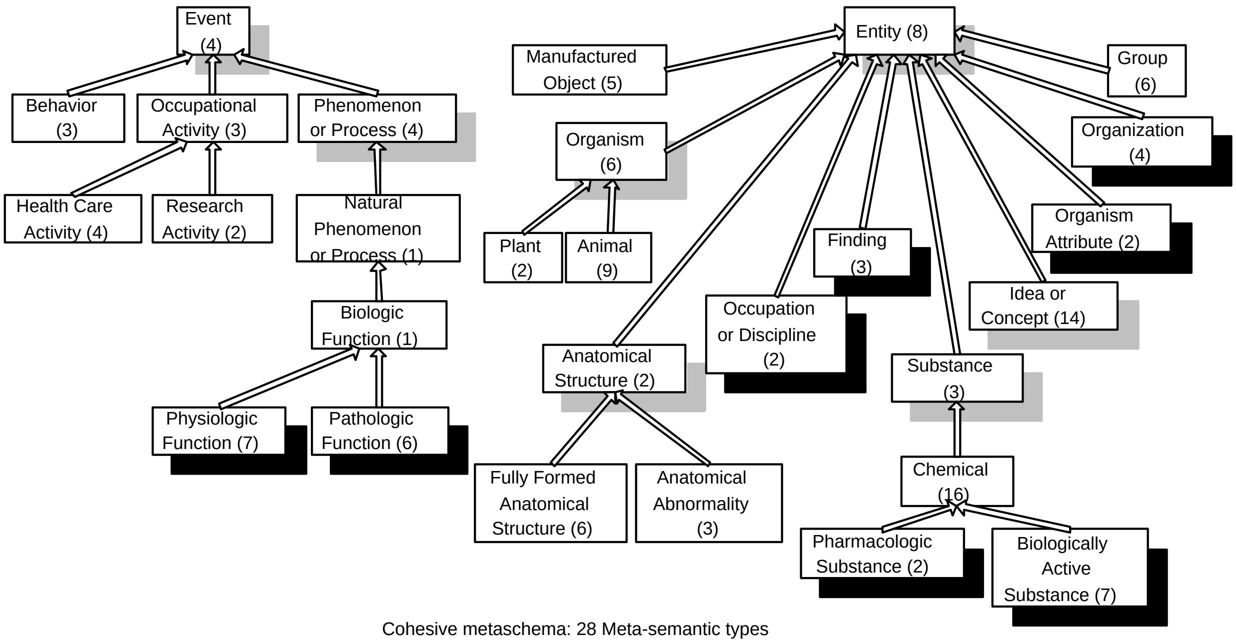
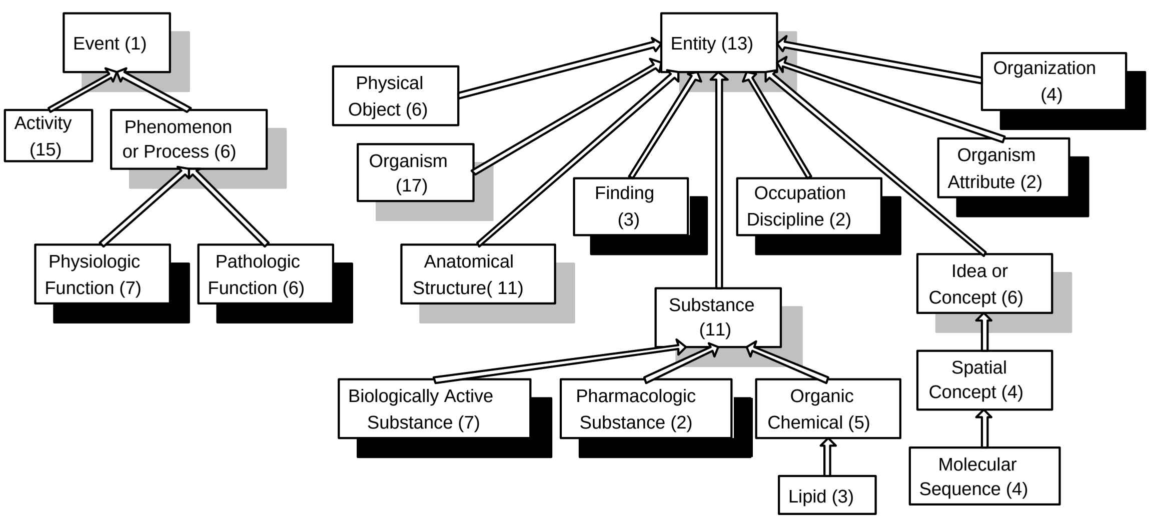


Figure 7: Entire lexical metaschema



Cohesive metaschema: 28 Meta-semantic types



Lexical metaschema: 21 Meta-semantic types

Figure 8: Hierarchies of cohesive and lexical metaschemas

types. Table 4 lists all the identical meta-semantic types and their sizes. Hence, both metaschemas agree that these eight meta-semantic types represent important subject areas in the SN. Altogether, they cover 33 semantic types (i.e., 24.4% of the SN).

Table 4: Identical meta-semantic types in the cohesive and lexical metaschemas

Meta-semantic type	Size
BIOLOGICALLY ACTIVE SUBSTANCE	7
FINDING	3
OCCUPATION OR DISCIPLINE	2
ORGANIZATION	4
ORGANISM ATTRIBUTE	2
PATHOLOGIC FUNCTION	6
PHARMACOLOGIC SUBSTANCE	2
PHYSIOLOGIC FUNCTION	7
Total: 8	33

In the following, we will write the number of semantic types (size) of a meta-semantic type in parentheses following its name for convenience of our comparison.

There are seven similar meta-semantic types. For example, PHENOMENON OR PROCESS(4) in the cohesive metaschema is similar to PHENOMENON OR PROCESS(6) in the lexical metaschema. Table 5 shows these similar meta-semantic types along with their sizes in each of the two metaschemas. In the cohesive metaschema, these seven meta-semantic types cover 41 semantic types, which is about 30.4% of the SN. In the lexical metaschema, these seven meta-semantic types cover 65 semantic types, which is about 48.1% of the SN.

To better understand the nature of the similarity represented in Table 5, we will explore refine-

Table 5: Similar meta-semantic types in cohesive and lexical metaschemas

Meta-semantic type	Size in cohesive metaschema	Size in lexical metaschema
ANATOMICAL STRUCTURE	2	11
EVENT	4	1
ENTITY	8	13
IDEA OR CONCEPT	14	6
ORGANISM	6	17
PHENOMENON OR PROCESS	4	6
SUBSTANCE	3	11
Total: 7	41	65

ments in both directions. As a refinement of the cohesive metaschema, consider the meta-semantic type IDEA OR CONCEPT(14) in the cohesive metaschema. This meta-semantic type is split into three separate meta-semantic types, IDEA OR CONCEPT(6), SPATIAL CONCEPT(4), and MOLECULAR SEQUENCE(4), in the lexical metaschema. In other words, {IDEA OR CONCEPT(6), SPATIAL CONCEPT(4), MOLECULAR SEQUENCE(4)} in the lexical metaschema is a refinement of IDEA OR CONCEPT(14) in the cohesive metaschema. This refinement case covers 14 semantic types in both metaschemas.

In the other direction, considering refinements of the lexical metaschema, there are three cases. As an example, {PHENOMENON OR PROCESS(4), NATURAL PHENOMENON OR PROCESS(1), BIOLOGIC FUNCTION(1)} in the cohesive metaschema is a refinement of PHENOMENON OR PROCESS(6) in the lexical metaschema. Table 6 shows these three cases of refinement of the lexical metaschema which cover 34 semantic types, in both metaschemas.

Table 6: Refinements in lexical metaschema

Meta-semantic type in lexical metaschema	Refinement in the cohesive metaschema
PHENOMENON OR PROCESS (6)	{PHENOMENON OR PROCESS (4), NATURAL PHENOMENON OR PROCESS (1), BIOLOGIC FUNCTION (1)}
ORGANISM (17)	{ORGANISM (6), PLANT (2), ANIMAL (9)}
ANATOMICAL STRUCTURE (11)	{ANATOMICAL STRUCTURE (2), FULLY FORMED ANATOMICAL STRUCTURE (6), ANATOMICAL ABNORMALITY 3}
Total: 3	34

Note that if there is a refinement case, then there is a meta-semantic type in one metaschema that is similar to one of the meta-semantic types in the refinement. For example, {PHENOMENON OR PROCESS(4), NATURAL PHENOMENON OR PROCESS(1), BIOLOGIC FUNCTION(1)} in the cohesive metaschema is a refinement of PHENOMENON OR PROCESS(6) in the lexical metaschema, where the PHENOMENON OR PROCESS meta-semantic types in both metaschemas are similar. However, not for every case of similar meta-semantic types is there a case of refinement. For example, ENTITY and EVENT are both cases of similarity, but they do not have refinements. The total number of semantic types covered by refinements in either direction is 48 (about 35.6%).

Besides the identical meta-semantic types, the similar meta-semantic types, and the meta-semantic types appearing in refinements, there are six meta-semantic types that appear exclusively in the cohesive metaschema; these are BEHAVIOR, OCCUPATIONAL ACTIVITY, HEALTH CARE ACTIVITY, RESEARCH ACTIVITY, MANUFACTURED OBJECT, and GROUP. There are four meta-semantic types that appear exclusively in the lexical metaschema; these are ACTIVITY, PHYSICAL OBJECT, ORGANIC CHEMICAL, and LIPID.

To summarize our comparison results, if we consider only the meta-semantic type names and not the underlying semantic-type groups, then 15 out of the 21 (about 71.4%) meta-semantic types in the lexical metaschema also appear in the cohesive metaschema. However, in the more precise comparison, there are 81 semantic types covered by identical meta-semantic types or refinements (about 60.0%). This shows that the cohesive metaschema and the lexical metaschema have only moderate similarity to each other.

4.3 Comparison between the Lexical Partition and the Semantic Partition

Since there is no metaschema for the semantic partition in [14], we can only compare our lexical partition to the semantic partition.

The lexical partition contains 21 semantic-type groups, while the semantic partition contains only 15 semantic-type groups. An obvious difference between these two partitions is that each group in the lexical partition is a connected group, rooted at a unique semantic type, while six groups in the semantic partition are not connected. Each disconnected group either contains isolated semantic types having no IS-A paths to any other semantic types in the group, or contains a forest with two or more trees.

There are only two identical semantic-type groups between the two partitions. They are the ORGANIZATION(4) group and the OCCUPATION OR DISCIPLINE(2) group. Altogether, they only cover six semantic types of the SN.

There is only one similar semantic-type group in the two partitions. The EVENT(1) semantic-type group in the lexical partition contains only one semantic type **Event**, which is a subset of the semantic-type group ACTIVITIES AND BEHAVIOR(9) in the semantic partition.

There is only one refinement case between the two partitions. {PHYSIOLOGIC FUNCTION(7), ORGANISM ATTRIBUTE(2)} in the lexical partition is a refinement for the PHYSIOLOGY(9) group in the semantic partition.

The total number of semantic types covered via either identical semantic-type groups or refinements is only 15 (11%). The large variation between the two partitions indicates that the two partitions are quite different from one another.

5 Discussion

The metaschema has been used for two applications. One of them is the creation of small views and the other is to audit the UMLS. The main use of metaschemas is in their support for comprehension of a large and complex network such as the UMLS SN. In [8, 11] we described how the framework of a metaschema supports the definition of a variety of partial views of the SN. By separately studying such partial views reflecting various subject areas of the SN, the user is able to slowly achieve orientation to parts of the SN, understand their interactions with one another, and eventually gain a comprehensive orientation to the whole complex SN. The orientation efforts may be supported by navigation of the metaschema. Such an approach takes into account the size limitation of both human mental capacity and graphical display devices. For details, see [8]. In [10] a metaschema is used to concentrate auditing efforts on a small number of concepts of the UMLS with high likelihood of errors. Those are concepts which are assigned to several semantic types belonging to different meta-semantic types. The high “semantic distance ” between semantic types of two different meta-semantic types raises the likelihood of an erroneous assignment for such concepts. All concepts associated with exactly the same set of semantic types form an intersection group. The results in [10] confirm this phenomenon for such small intersection groups of less than ten concepts. For details see [10].

The lexical technique was one of two techniques we used in [13] to identify extra IS-A relationships to obtain the Enriched Semantic Network (ESN) of the UMLS. The ESN hierarchy has a directed acyclic graph (DAG) structure and consists of 139 semantic types and 150 IS-A relationships. Out of the 19 new IS-A relationships of the ESN, which are not in the SN, 12 define lexically related CP-pairs (63%), compared to 70% in the original SN, as mentioned before.

We compared the new *lexical metaschema* with our previously derived *cohesive metaschema*. The two metaschemas are shown to be moderately similar. A natural question is: which of the two would be more appropriate to support user orientation. Reviewing the two metaschemas and specially their differences (See Figure 8), we see that each of them has strengths and weaknesses.

For example, we find that the meta-semantic type PHENOMENON OR PROCESS of six semantic types in the lexical metaschema has a refinement into the three meta-semantic types {PHENOMENON OR PROCESS(4), NATURAL PHENOMENON OR PROCESS(1), BIOLOGIC FUNCTION(1)} in the cohesive metaschema. We prefer the one set of six, as unless essential, we do not like singleton in the metaschema. In another case, we prefer the refinement {IDEA OR CONCEPT(6), SPATIAL CONCEPT(4), MOLECULAR SEQUENCE(4)} of the lexical metaschema to the meta-semantic type IDEA OR CONCEPT(14) in the cohesive metaschema, since the corresponding group in the cohesive partition is too large, too wide and non-homogeneous, in its coverage. The partition into these three medium-sized meta-semantic types in the lexical metaschema is into meaningful subject areas and is preferred. But there are examples where we prefer the cohesive metaschema. For example, we prefer the meta-semantic type MANUFACTURED OBJECT(5) of the cohesive metaschema to the corresponding meta-semantic type PHYSICAL OBJECT(6) in the lexical metaschema which contains the extra semantic type **Physical Object**. In our opinion, **Physical Object** is a more general terms which belongs to the ENTITY meta-semantic type which contains other general terms, while the semantic types in the meta-semantic type MANUFACTURED OBJECT are more specific terms.

In conclusion, for both metaschemas obtained algorithmically, many groups obtained seem semantically proper, but some groups seem improper. Each of the techniques fails in different subject areas.

In a stark contrast, the comparison between the lexical partition and the semantic partition shows very little resemblance. The main reason is that these two partitions do not follow the same basic principles. The lexical partition is into connected groups, disallowing singleton leaves. The groups are named after their root semantic types.

The semantic partition groups are chosen to cover subject areas in the field of medicine and do not necessarily correspond to semantic types. Furthermore, the groups are not required to be connected. Due to this last property, the semantic partition does not fit for defining a metaschema. As a matter of fact, nine out of the 15 groups are not connected. Due to this flexibility, the groups of the semantic partition are different from the lexical groups containing some of the same semantic types. For example, consider the ANATOMY group of the semantic partition versus the ANATOMICAL STRUCTURE group of the lexical partition. Both groups have eleven semantic types, but the ANATOMY group contains four isolated semantic types which are not connected to one another: **Body Substance**, **Body Location or Region**, **Body Space or Junction**, and **Body System**. On the other hand, the group is missing four semantic types which belong to the ANATOMICAL STRUCTURE group of the lexical partition, namely, **Gene or Genom** which was put in the GENE AND MOLECULAR SEQUENCES group in the semantic partition, and **Anatomical Abnormality** with its two children which were put in the DISORDERS group in the semantic partition. We also note that in the evaluation of the semantic groups by visual approaches in [15] by the same authors, the conclusion of their evaluation is that some of their groups should be changed. For example, the disconnected LIVING BEINGS group should be split into two connected groups ORGANISM and GROUP, the first of which matches the ORGANISM group of the lexical partition. In conclusion, taking into account the different nature and assumptions of these two partitions, it is not surprising that they are so different.

6 Conclusions and Future Work

A metaschema is a compact, abstract view of the SN. Various metaschemas are possible. In this paper, we presented the lexical metaschema derived via an algorithmic lexical partitioning approach. In previous work [8], we presented the cohesive metaschema derived according to purely structural considerations. In that metaschema, each meta-semantic type represented a group of semantic types with the same (or almost the same) relationships. These two algorithmically derived metaschemas were compared in this paper and we found them to be moderately similar. Depending on the comparison method used, we found a similarity value of 71.4% from meta-semantic level or 60% from semantic type coverage level. A natural question is: which of these two metaschemas is better in supporting user orientation to the SN? To answer this question, we need a way to measure the overall quality of a given metaschema. As can be expected, each metaschema has its advantages and disadvantages. This observation leads us naturally to ask: can we construct a metaschema that incorporates the “good parts” of each of the above metaschemas while avoiding their pitfalls? These issues will be addressed in our future research.

Acknowledgement

We thank the anonymous reviewers for their insightful remarks which helped to enhance this work.

References

[1] Campbell KE, Oliver DE, and Shortliffe EH. The Unified Medical Language System: Toward a collaborative approach for solving terminologic problems. *JAMIA*, 5(1):12–16, 1998.

[2] Humphreys BL, Lindberg DAB, Schoolman HM, and Barnett GO. The Unified Medical Language System: An informatics research collaboration. *JAMIA*, 5(1):1–11, 1998.

[3] Lindberg DAB, Humphreys BL, and McCray AT. The Unified Medical Language System. *Methods of Information in Medicine*, 32:281–291, 1993.

[4] Schuyler PL, Hole WT, Tuttle MS, and Sherertz DD. The UMLS Metathesaurus: Representing different views of biomedical concepts. *Bull Med Libr Assoc*, 81(2):217–222, 1993.

[5] Tuttle MS, Sherertz DD, Olson NE, Erlbaum MS, Sperzel WD, and Fuller LF et al. Using META-1, the first version of the UMLS Metathesaurus. In *Proc. Fourteenth Annual SCAMC*, pages 131–135, 1990.

[6] McCray AT. Representing biomedical knowledge in the UMLS Semantic Network. In Broering NC, editor, *High-Performance Medical Libraries: Advances in Information Management for the Virtual Era*, pages 45–55, Mekler, Westport, CT, 1993.

[7] McCray AT and Hole WT. The scope and structure of the first version of the UMLS Semantic Network. In *Proc. Fourteenth Annual SCAMC*, pages 126–130, Los Alamitos, CA, November 1990.

[8] Perl Y, Chen Z, Halper M, Geller J, Zhang L, and Peng Y. The cohesive metaschema: a higher-level abstraction of the UMLS Semantic Network. *Journal of Biomedical Informatics*, 35(3):194 – 212, 2003.

[9] Chen Z, Perl Y, Halper M, Geller J, and Gu H. Partitioning the UMLS Semantic Network. *IEEE Trans. Information Technology in Biomedicine*, 6(2):102–108, June 2002.

[10] Gu H, Min H, Peng Y, Zhang L, and Perl Y. Using the metaschema to audit UMLS classification errors. In *Proc. 2002 AMIA Annual Symposium*, pages 310–314, San Antonio, TX, November 2002.

[11] Zhang L, Perl Y, Halper M, and Geller J. Designing metaschemas for the UMLS Enriched Semantic Network. *Journal of Biomedical Informatics*, 36(6):433–449, Dec 2003.

[12] Zhang L, Perl Y, Geller J, Halper M, and Cimino JJ. Enriching the structure of the UMLS Semantic Network. In *Proc. of the 2002 AMIA Annual Symposium*, pages 939–943, San Antonio, TX, November 2002.

[13] Zhang L, Perl Y, Geller J, Halper M, and Cimino JJ. An enriched UMLS Semantic Network with a multiple inheritance hierarchy. To appear in JAMIA.

[14] McCray AT, Burgun A, and Bodenreider O. Aggregating UMLS semantic types for reducing conceptual complexity. In *Proc. Medinfo 2001*, pages 171–175, London, UK, September 2001.

[15] Bodenreider O and McCray AT. Exploring semantic groups through visual approaches. *Journal of Biomedical Informatics*, 36(6):414–432, December 2003.

[16] Nelson SJ, Sheretz DD, Tuttle MS, and Erlbaum MS. Using MetaCard: A hypercard browser for biomedical knowledge sources. In *Proc. Annu Symp Comput Appl Med Care*, pages 151–154, 1990.

[17] Li Q, Shilane P, Noy NF, and Musen MA. Ontology acquisition from on-line knowledge sources. In *Proc. 2000 AMIA Annual Symposium*, pages 497–501, Los Angeles, CA, November 2000.

[18] Tuttle MS, Cole WG, Sheretz DD, and Nelson SJ. Navigating to knowledge. *Methods Inf Med*, 34(1-2):214–231, 1995.

[19] Aristotle. *The categories*. Cambridge, MA: Harvard University Press, 1973.

[20] Han J and Kamber M. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, August 2000.