Using Semantic Links for Information Extraction and Semantic Representation

Kerstin Denecke

University of Hannover, Research Center L3S, Appelstr. 9a, D-30167 Hannover
denecke@l3s.de

Abstract. Ontologies and possibilities of using and enhancing them is a currently widespread research area. Widely unused are relations between concepts defined in ontologies in the context of knowledge representation. In this paper, the potential of using a semantic network to support the automatic generation of semantic structures is analysed. Semantic representations to unstructured natural language documents are generated by means of the method SeReMeD. This method maps a natural language document to concepts of the Unified Medical Language System (UMLS). Contextual relations expressed in natural language are automatically identified and represented in the generated structures. To obtain additional semantic relationships, the UMLS Semantic Network and relationships between concepts predefined in the UMLS Metathesaurus are used to support the structuring process. By means of these relations, automatically generated semantic structures can be enhanced and ameliorated.

Keywords: Natural Language Understanding, Natural Language Processing, Ontology

1 Introduction

For some time, the attempts to use natural language processing (NLP) techniques in the field of medicine increase. They aim to facilitate the access to data in unstructured medical documents and to allow the re-use of data by a couple of applications such as quality assessment or decision support. NLP techniques have demonstrated the potential to unlock information from medical texts. Using these methods natural language text is transformed into standardised and normalised semantic structures. From these structures, key information can be extracted and be made available for further applications. Hence, the expenditure of time and money for documentation and retrieval purposes can be reduced particularly by reusing existing data.

Ontologies supply vocabularies and world knowledge necessary for clear communication within a certain domain. In the domain of medicine, ontologies supply words or word groups (concepts) that are used in the domain of medicine. Often, they provide information on relationships between the concepts (e.g. in SNOMED CT there is an ‘isa’ relationship between the concepts ‘appendix’ and ‘large intestine part’).

Existing NLP systems, like MedLEE, HITEx or LUPUS, normally use medical ontologies only for mapping natural language to corresponding concepts of the terminology. In MedLEE (Medical Language Extraction and Encoding System, developed by Friedman et al., [10]), a system for extracting and coding data from radiology and pathology reports, natural language is mapped to concepts of the Medical Entries Dictionary (MED) or of the UMLS [16], respectively.

Zeng et al. [5] developed an open source medical NLP system (HITEx). Several existing language engineering technologies are used by that system, such as GATE and several of its NLP modules or WEKA for classifying sentences. The UMLS provides the medical knowledge resource: In HITEx, natural language strings are mapped to UMLS concepts.

The Lung Pathology System (LUPUS, [7]) aims to extract data from medical pathology reports in German. LUPUS combines a NLP module with a Semantic Web Component which defines the domain ontology and is specified in OWL. LUPUS generates a semantic representation using a linguistic analysis. The ontology provides the vocabulary for the semantic representation and supports the resolution of ambiguities. A more comprehensive overview of current approaches in medical NLP is given by Spyns [11], Hobbs [13] and Hersh [12].

The mentioned systems use concepts from ontologies, but relationships provided by an ontology are widely unused in medical NLP systems. Baclawski et al. describe in [1] an approach to exploit the UMLS for generating and indexing knowledge representations for medical documents. Their representation rules are based on syntactic information and semantic knowledge from the ontology, respectively. In addition, the terminology is used for assigning terms to concepts and categories of the terminology.

MedIE [2], developed at Drexel University, aims to extract terms and relations from English medical text. This system determines terms that are syntactically or semantically related to each other. A Link Grammar Parser is used to find syntactically relations, while relations predefined in the terminology (UMLS) help to find semantically relations. This approach tries to combine existing language engineering technologies such as WordNet as lexical-semantic resource and GATE for tokenising and tagging.
Nevertheless, the described approaches do not aim to generate a semantic representation for a complete document. Therefore, this paper analyses the potentials of using relationships predefined in medical terminologies. In particular, it will be analysed whether relations defined in the UMLS or in the UMLS Medical Semantic Network (UMLS MSN) can help to avoid errors in structure generation using SeReMeD [15]. The objective is to produce more correct semantic structures for a medical document and to ameliorate and enrich the semantic structures generated by SeReMeD.

2 The Unified Medical Language System

The UMLS (http://umlsinfo.nlm.nih.gov/) combines many medical vocabularies and provides a mapping structure between them. It is composed of three main knowledge components: the Metathesaurus, the Semantic Network and the SPECIALIST Lexicon. The Metathesaurus integrates vocabularies from the biomedical domain (e.g. MeSH, SNOMED CT). Each concept of the Metathesaurus has specific attributes that define its meaning. Concepts are assigned to at least one semantic type of the UMLS Medical Semantic Network (MSN, see below). The biomedical vocabularies contribute thesaural relationships between the concepts (e.g., 'is-a' - relationships).

The 'is-a' - relation is the basic hierarchical link in the Metathesaurus. If two items are linked by an 'is-a' - relation, the first item is more specific in meaning than the second item. Normally, an 'is-a' relation only exists between semantic concepts belonging to the same semantic type.

The UMLS MSN [8] is a network of general semantic categories or types. It provides 134 semantic types that are grouped in turn 15 semantic groups (e.g. the concept atrial fibrillation belongs to the semantic types Finding and Pathologic Function that in turn belong to the group Disorders). The MSN defines semantic relationships between the semantic types. It provides hierarchical relationships ('is-a') and non-hierarchical relations ("physically related to", "spatially related to", "temporally related to", "functionally related to", "conceptually related to").

The SPECIALIST Lexicon provides the linguistic knowledge that comprises syntactical information on terms as well as natural language processing tools (e.g. tokenizer).

3 SeReMeD – Method for generating semantic structures

SeReMeD allows for the automatic generation of semantic representations. Out of an unstructured medical document, relevant concepts are identified and linked to each other according to the semantic expressed in the document. For this purpose, a processing pipeline consisting of nine processing steps is used. These steps are described in the following sections and illustrated in Fig. 1.

3.1 Section Splitter

At the beginning of the processing pipeline, the Section Splitter dissects a document into several subsections. Each section is assigned to one of the section classes that are defined for the different document types using regular expressions (e.g., the section class admission in a discharge summary or the section class finding in a radiology report).

3.2 Sentence Splitter

Then, each section is splitted into sentences using a preparser. Based on a shallow syntactic analysis provided by a simple parser, each sentence is decomposed into segments using prepositions. A segment is a smaller part of a sentence that reaches from the beginning of a sentence to the first preposition (head segment) or it starts with a preposition and reaches to the next preposition or to the end of the sentence (prepositional segment). E.g., the sentence Heart size and vessels are within normal limits is splitted into the head segment Heart size and vessels are and the prepositional segment within normal limits.

3.3 Sentence Tokenizer

The Sentence Tokenizer splits each sentence into tokens (substrings). The tokens are separated by delimiters (e.g. blanks, punctuation).
3.4 POS Tagger

Then, the POS Tagger classifies each token, i.e. assigns tokens to a corresponding word class (part of speech). Its morphologic and linguistic properties are determined (e.g. its morphemes, gender, case and number).

3.5 Negation Detector and Special Expression Finder

After these preprocessing steps, each sentence is checked for special expressions by means of regular expressions. Special expressions include dates, quantities and dosage specifications, expressions with special meanings, like “exclusion of” or “evidence of”. All these special expressions are tagged before the indexing process starts and are interpreted separately. Furthermore, negated structures like No effusions are identified are identified using regular expressions. Concepts that are negated in the input sentence are represented as negated concepts in the semantic structure.

3.6 Concept Mapper

Next, the token set of each segment is mapped to a set of one or more concepts of the underlying medical terminology (UMLS) by the Concept Mapper. The Concept Mapper uses MetaMap to find propositions of UMLS concepts for a sentence. MetaMap [3] is based on NLP techniques and maps natural language to concepts of the UMLS. It provides a ranked set of candidates for concepts that are probably expressed by an input phrase. Out of the possible candidates provided by MetaMap, the Concept Mapper selects the first proposal out of the highest ranked candidate set.
3.7 Concept Generator

For each selected UMLS concept, a semantic concept is created that consists of the free-textual description of the UMLS concept, a semantic type (which is the semantic type of the UMLS concept), a unique identifier (which is the UMLS Concept ID), a main category (which is the semantic group to which the semantic type belongs) and optionally, a semantic role (e.g., NEG, IDENTIFIED).

3.8 Semantic Interpreter

Out of the set of semantic concepts, provided by the Concept Generator, the Semantic Interpreter determines for each segment a main information unit (leading concept). All other information is considered as modifying information (modifying concepts). The main information is ascertained using the assumption that each sentence deals with one semantic entity, which in clinical narratives is normally a diagnosis / morphological change or a procedure / treatment. Concepts describing a diagnosis or morphological change belong to the UMLS semantic categories Disorder, Physiology or Phenomena whereas the category Procedure contains concepts on procedures and treatments. Concepts of one of these semantic categories are considered as central information units which are specified by all other concepts.

The mentioned assumption is reflected by manually arranged priority sequences of UMLS semantic groups. The concept whose main category is in the furthest left position in a priority sequence, will be chosen as main information.

3.9 Structure Generator

The Structure Generator introduces relations between concepts: For each segment, the modifying concepts are linked to the main information. The relation type of the corresponding relations is predefined by the semantic type to which the modifying concept belongs. Furthermore, the main information of the prepositional segments is linked to the main information of the head segment. It is assumed that the information described in prepositional segments modifies the information of the head segment. The relation type of the introduced relation is given by the meaning of the preposition, which is interpreted according to the context it is used in (e.g., the preposition “with” can express the meaning using or accompanied by).

Using the described procedure, each sentence is mapped to a conceptual graph-like semantic representation of its relevant medical contents (see example 1).

Example 1: Semantic structures generated by SeReMeD

(1) Cardiomegaly without failure.
   (Finding) cardiomegaly
   |- (NEG) (Functional Concept) Failure  
   The negated structure is represented by the semantic role NEG.

(2) Partial atelectasis of the right lower lobe.
   (Pathologic Function) Atelectasis
   |- (Quantitative Concept) Partial
   |- (topology) (Body Part, Organ, Organ Component) Structure of right lower lobe of lung
   The relation type topology is derived from the meaning of the preposition of in the given context.

SeReMeD is described in more detail in [15]. Compared to the method described there, the processing pipeline mentioned above is a modified version that processes documents in English and uses the MetaMap Algorithm for mapping strings to UMLS concepts.

An evaluation in [15] showed that SeReMeD produces semantic representations of high quality for phrases with small complexity. But for specific syntactic-semantic structures (e.g., coordinated structures, sentences with coreferences), it is still not possible to correctly identify the main and modifying information. In addition, SeReMeD is only able to determine dependencies between main information and modifying information. Relations between modifying information or main information itself are not considered (see example 2). For this reason, it will be analysed, whether semantic relationships of a semantic network can provide additional information for avoiding these errors.

Example 2

For the phrase Inconspicuous representation of the pancreas organ SeReMeD generates the following semantic structure:
4 Using predefined semantic relations for generating semantic structures

In the following, it will be shown how to support the generation process of SeReMeD by relations predefined in a medical terminology and how to integrate additional knowledge in the semantic structure in this way. The aim is to generate more detailed semantic structures and to identify dependencies hidden for SeReMeD.

Two different kinds of relations are available while using UMLS as knowledge resource: The relationships between semantic concepts predefined in the Metathesaurus and the relationships between semantic types provided by the Medical Semantic Network. They offer different possibilities to support SeReMeD. Their usage is therefore described separately in the following sections. Furthermore, the integration of the proposed method for using semantic relations into SeReMeD will be shown.

4.1 Usage of relations between semantic concepts

In this section, the usage of relations defined in the Metathesaurus for structure generation is introduced. More precisely, it will be analysed how to use the “is-a”-relation in this context.

Sometimes, several concepts of the same semantic category are selected as main information by the Semantic Interpreter. But mostly, one of these concepts modifies the other concept(s). Using hierarchical relations, it can be checked whether a concept that is selected as main information is subsumed of another concept that is determined as main information. In this way, also dependencies between modifying concepts can be determined.

For using is-a-relationships during structure generation, Rule 1 is considered while generating semantic representations.

**Rule 1:**
If the terminology defines an “is-a” - relation between two concepts selected as leading concepts or between two modifying concepts of the same segment, the semantic structure will only contain the more specific concept. The generic term will be excluded from representation.

Considering Rule 1, is-a-relationships can help to exclude redundant concepts of being represented. Example 3 illustrates the usage of this rule.

**Example 3**

(1) Consider the phrase *Inconspicuous representation of the pancreas organ* that was already mentioned in Example 2. In the Metathesaurus the following relations are defined:

*Pancreas* ➔ *isA* ➔ *digestive organ* ➔ *isA* ➔ *Body organ*

Using Rule 1, the concept *[Body Organ]* is excluded from representation because it does not provide any additional information:

(Conceptual Entity) Representation Component
- (topology) (Body Part, Organ, or Organ Component) Entire pancreas
- (topology) (Body Part, Organ, or Organ Component) Body Organ

(2) The semantic structure of the phrase *Old right fracture (rib fracture)* contains a redundant information: The method SeReMeD determines *[Injury or Poisoning: Fracture]* and *[Injury or Poisoning: Rib Fracture]* as main information and the following semantic structure is provided:

(Injury or Poisoning) Rib Fracture
- (Temporal Concept) Old episode
- (Spatial Concept) Right

(Injury or Poisoning) Fracture
- (Temporal Concept) Old episode
- (Spatial Concept) Right
The Metathesaurus defines the relations:
Fracture of rib $\rightarrow$ isA $\rightarrow$ Fracture of bones of trunk $\rightarrow$ isA $\rightarrow$ Fracture

Considering Rule 1 and these relations, a modified, more compact semantic structure can be generated:

\[
\text{(Injury or Poisoning) Rib Fracture} \\
|\text{- (Temporal Concept) Old episode} \\
|\text{- (Spatial Concept) Right}
\]

4.2 Usage of relations between semantic types

The UMLS Medical Semantic Network defines relationships among semantic categories. Using UMLS MSN relationships, dependencies between concepts of one segment can be identified. For this purpose, MSN relations must be considered by the Structure Generator using Rule 2.

**Rule 2:**

A MSN relation between the semantic types of two concepts of the same segment should be represented by the semantic structure: The concepts are linked to each other according to the relation.

The application of this rule is demonstrated in Example 4. In this way, the MSN relationships can help to enrich the semantic structures and to produce more detailed structures.

**Example 4:**

(1) For the sentence *There has been interval increase in the left upper lobe air space disease* SeReMeD generates the following semantic structure:

\[
\text{(Idea or Concept) Data Type Interval} \\
|\text{- (localisation) (Body Part, Organ, Organ Component) Structure of left upper lobe of lung} \\
|\text{- (Disease or Syndrome) Disease} \\
|\text{(Functional Concept) Increase} \\
|\text{- (localisation) (Body Part, Organ, Organ Component) Structure of left upper lobe of lung} \\
|\text{- (Disease or Syndrome) Disease}
\]

In this structure, the relation between the two concepts \[\text{Idea or Concept: Data Type Interval}\] and \[\text{Functional Concept: Increase}\] is missing. Instead, both are represented as main concepts. The UMLS MSN provides a relation between the semantic types of the two concepts, that is

\[
\text{Functional Concept} \rightarrow \text{isa} \rightarrow \text{Idea or Concept}
\]

Using Rule 2, the following semantic structure can be generated using this relation:

\[
\text{(Idea or Concept) Data Type Interval} \\
|\text{- (Functional Concept) Increase} \\
|\text{- (localisation) (Body Part, Organ, Organ Component) Structure of left upper lobe of lung} \\
|\text{- (Disease or Syndrome) Disease}
\]

In this way, the relationship helps to generate dependencies between two main information concepts.

(2) For generating a semantic structure for the sentence *Small bilateral nodules are again noted, unchanged* the system generates the semantic structure:

\[
\text{(Acquired Abnormality) Nodule} \\
|\text{- (Quantitative Concept) Small} \\
|\text{- (Spatial Concept) Bilateral} \\
|\text{(Intellectual Product) NOTE} \\
\text{(Finding) No status change} \\
|\text{- (Quantitative Concept) Small} \\
|\text{- (Spatial Concept) Bilateral} \\
|\text{(Intellectual Product) NOTE}
\]

The representation is missing the link between the two main information concepts, namely that \[\text{Acquired}\]
Abnormality: Nodule is modified by [Finding: No status change]. The UMLS MSN provides the relationship Acquired Abnormality \( \Rightarrow \) associated with \( \Rightarrow \) Finding.

This relation can be integrated into the semantic structure (Rule 2) and leads to the semantic structure as follows.

\[
\begin{align*}
\text{Acquired Abnormality} \quad & \text{Nodule} \\
\Rightarrow & \text{associated with} \quad \text{Finding} \quad \text{No status change} \\
\Rightarrow & \text{Quantitative Concept} \quad \text{Small} \\
\Rightarrow & \text{Spatial Concept} \quad \text{Bilateral} \\
\Rightarrow & \text{Note}
\end{align*}
\]

(3) The MSN relationship can also help to structure modifying concepts. In the semantic structure for the sentence Further evaluation with a computed tomography-scan is suggested:

\[
\begin{align*}
\text{(Health Care Activity) Assessment procedure} \\
\Rightarrow & \text{(Spatial Concept) Further} \\
\Rightarrow & \text{(using) (Diagnostic Procedure) X-Ray CT} \\
\Rightarrow & \text{(using) (Diagnostic Procedure) Scanning}
\end{align*}
\]

two concepts of the semantic category Diagnostic Procedure occur. Considering the relation Diagnostic Procedure \( \Rightarrow \) method_of \( \Rightarrow \) Diagnostic Procedure a more detailed semantic structure can be provided.

\[
\begin{align*}
\text{(Health Care Activity) Assessment procedure} \\
\Rightarrow & \text{(Spatial Concept) Further} \\
\Rightarrow & \text{(using) (Diagnostic Procedure) Scanning} \\
\Rightarrow & \text{(method_of) (Diagnostic Procedure) X-Ray CT}
\end{align*}
\]

The mentioned examples give an insight into the usage of a semantic network for generating semantic structures. They suggest that the relations can help to produce more detailed semantic structures and to exclude redundant information from representation.

### 4.3 Integration into SeReMeD

For using semantic relations as described above in SeReMeD, a new processing module is introduced. Based on the information on main and modifying concepts that are provided by the Semantic Interpreter, the new module (called Relation Finder) searches for ontology relations. It searches for UMLS Metathesaurus relationships between concepts of the same semantic category. In addition, the Relation Finder tries to determine UMLS MSN relationships between semantic types (see pseudo code below) of concepts. Taking into account the information of the Semantic Interpreter as well as the relations found by the Relation Finder, the Structure Generator then links the corresponding concepts and generates the semantic representation.

**Fig. 2.** Pseudo code for determining relations in SeReMeD

```plaintext
For each phrase segment{
    For each pair of concepts (i, j) i, j \( \in \) \( C_{\text{main}} \) or i, j \( \in \) \( C_{\text{mod}} \) with same semantic category{
        Search for UMLS Metathesaurus relation R(i, j) of type ‘is-a’
        If (R(i, j) exists) exclude j from representation
    }
    For each pair of concepts (i, j) i, j \( \in \) \( C_{\text{main}} \) or i, j \( \in \) \( C_{\text{mod}} \) {
        Search for UMLS MSN relations R(type\(_i\), type\(_j\))
        If (R(i, j) exists) link i to j in the semantic structure using
        the relation of type R(type\(_i\), type\(_j\))
    }
}

\( C_{\text{mod}} \) – set of modifying concept, \( C_{\text{main}} \) – set of main concept
```
5 Experiments

Instead of performing a complete evaluation, some experiments have been started first. The main goal of the described experiments is to estimate the usefulness of the mentioned extensions of SeReMeD. The question whether considering ‘external’ semantic relations can help to produce high quality semantic representations is addressed.

To achieve the evaluation dataset, 50 radiology reports were represented semantically by SeReMeD. Out of the generated structures, 50 sentences, whose semantic structures contain more than one main information or more than one modifying information for at least one segment had been selected.

In the semantic structures of half of these selected sentences, the problem of having more than one main information occurred. The structures of four sentences missed dependencies between modifying information, 15 sentences has been correctly represented. For six sentences, other kinds of error lead to incorrect semantic structures (see table 1).

Table 1: Sources of error in the selected sentences

<table>
<thead>
<tr>
<th>Source of Error</th>
<th>Number of sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>More than one main information</td>
<td>25</td>
</tr>
<tr>
<td>Missing dependencies between modifying concepts</td>
<td>4</td>
</tr>
<tr>
<td>Others</td>
<td>6</td>
</tr>
</tbody>
</table>

For 25 sentences with wrong semantic structures, relations between semantic types of either the main or the modifying information could be identified in the UMLMSN. Normally, more than one relation between two semantic types is provided by the UMLMSN. Assuming that always the correct relation is selected, SeReMeD could produce correct semantic structures.

The first results from the experiments suggest that the problem of missing relations in the semantic structures produced by SeReMeD occurs quite often. Mostly, these missing relations can be gained by means of UMLSS relationships. A more comprehensive statement regarding possible improvements can be provided as soon as a complete evaluation with a larger data set has been performed.

6 Discussion

The UMLSS Metathesaurus is used by different natural language processing systems. Friedman et al. modify their MedLEE system and try to use UMLSS concepts [16]. In HITEx [5], the UMLSS Metathesaurus provides the domain knowledge of the systems. All of these approaches are not using the UMLSS Medical Semantic Network relations during processing. In MedIE [2], UMLSS relationships of the Metathesaurus are used to extract relations between terms (concepts) in a document. In this paper, an approach has been introduced to use UMLSS relations both predefined in the Metathesaurus and in the UMLSS MSN for knowledge representation. It has been shown how to use these relations for generating and enrich semantic structures.

The method SemRep, introduced by Rindflesh and Aronson in [17] is comparable to the one presented in this paper. SemRep uses two types of semantic rules to interpret phrases semantically: Linguistic patterns are matched against corresponding relationships between UMLSS MSN semantic types (e.g., the preposition “in” corresponds to the MSN relation “part-of”). The other kind of rules depends on the UMLSS semantic types [17]. But in contrast to the approach here, SemRep aims not to determine a semantic structure for a complete sentence. It provides semantic dependencies between concepts of an input phrase without mapping all relevant information of a sentence to a semantic structure and creating one single representation for a sentence (see example 5).

Example 5: Processing result of SemRep

Considering the sentence Right lateral decubitus view of the chest SemRep [17] only extracts the relation

Anterior thoracic region LOCATION_OF Decubitus ulcer

whereas SeReMeD provides a semantic structure for the whole phrase:

(Organism Attribute) View
    (Spatial Concept) Lateral to the right
    (Spatial Concept) Decubitus direction
    (topology) (Body Location or Region) Anterior thoracic region

Nevertheless, there are some aspects in the approach of Aronson and Rindflesch that could be used within SeReMeD: The semantic rules proposed there could probably help to determine dependencies between prepositional segments.
One problem occurs while querying the relations of the Medical Semantic Network: Sometimes, there exist several relationships between two semantic types. It is unclear, which one has to be selected for enriching the semantic structure of a sentence correctly (see Example 7). For this purpose, the context of the concepts needs to be considered.

Example 6

The sentence *Advanced degenerative disease both the humeral scapular joints and the acromioclavicular joints* is represented as follows:

(Finding) advanced disease
- (Body Space or Junction) Joint structure of shoulder region
- (Body Part, Organ, or Organ Component) Bone structure of scapula
- (Body Space or Junction) Acromioclavicular joint structure

(Pathologic Function) Degenerative abnormality
- (Body Space or Junction) Joint structure of shoulder region
- (Body Part, Organ, or Organ Component) Bone structure of scapula
- (Body Space or Junction) Acromioclavicular joint structure

Taking the UMLS MSN relations into consideration, the concept *[Pathologic Function: Degenerative abnormality]* should be represented as modifying the concept *[Finding: advanced disease]*. But the relation type cannot be determined, because there are three different relationships defined that link the semantic types Finding and Pathologic Function, namely:

- Finding ➔ associated_with ➔ Pathologic Function
- Finding ➔ evaluation_of ➔ Pathologic Function
- Finding ➔ manifestation_of ➔ Pathologic Function

Vintar *et al.* address this problem in [20] by an approach for relation filtering with inverse document frequency or verbal markers that also may be helpful in the context analysed in this paper. Relations introduced by verbs are widely unused in SeReMeD. In [18], *Schutz and Buitelaar* describe a method for extracting relations from text. They propose to use the extracted relations for extending an ontology. In the approach here, verb relationships could additionally enhance the generated semantic structures.

Another difficulty with the UMLS MSN arises for the following reason: Given a relationship between two semantic types, two concepts belonging to these types are not necessarily linked by this relationship. So, if the relation between the semantic types is used to enrich the semantic structure, possibly two concepts are linked that are not semantically correlating.

Nevertheless, using semantic relationships for structure generation (i.e. knowledge representation) provides several advantages: The resulting semantic structure can be more correctly by representing more relations expressed in a textual document. Relationships that are implicitly contained in a sentence can be represented as well. These relations are accessible using world knowledge defined in an ontology. Furthermore, redundant information can be excluded from representation.

7 Conclusion

In this paper, it has been analysed how to use semantic relations predefined in the UMLS and the UMLS Medical Semantic Network. It can be noticed, that these relationships can be considered during structure generation with SeReMeD. They can help to exclude redundant information from representation and to generate more detailed knowledge representations. In this way, they can support the structuring process of SeReMeD especially for handling more complex sentence structures.

Future research will include the filtering of UMLS relations between semantic types according to the context to tackle the problems mentioned above. In addition, an evaluation has to be performed to be able to give more meaningful statements regarding the possible improvements of the proposed idea. Another future research direction will be the usage of a semantic network for supporting the extraction of specific information out of the semantic structures. In [19], an approach for ontology-based information extraction is already presented. In that work, an ontology is used to interpret and link extracted information to concepts of the ontology.
References


