

# Modellierung und Integration von Wortnetzen und Domänenontologien in OWL am Beispiel von GermaNet und TermNet

**Claudia Kunze**

Ruprecht-Karls-Universität Heidelberg

**Lothar Lemnitzer**

Eberhard Karls Universität Tübingen

**Harald Lünzen**

Justus-Liebig-Universität Gießen

**Angelika Storrer**

Universität Dortmund

## Abstract

Der Beitrag beschreibt ein Konzept zur Modellierung des allgemeinsprachlichen Wortnetzes GermaNet und des domänenspezifischen Wortnetzes TermNet in der Web Ontology Language OWL. Während sich die Modellierung für das allgemeinsprachliche Wortnetz an entsprechenden Vorschlägen zum englischen Princeton WordNet orientiert, wird für das fachsprachliche Wortnetz ein alternatives Modellierungskonzept entwickelt, das den Besonderheiten domänenspezifischer Terminologien Rechnung trägt. Wir stellen einen Vorschlag zur Verknüpfung von allgemeinsprachlichem und terminologischem Wortnetz im Rahmen von OWL vor und diskutieren auf dieser Basis Probleme und Modellierungsalternativen.

## 1 Einführung

Wortnetze haben vielfältige Anwendungsfelder in der Sprach- und Texttechnologie sowie im Information Retrieval, vgl. z.B. (?; Kunze et al.(2003)Kunze, Lemnitzer und Wagner). In aktuellen Forschungen werden zunehmend Vorschläge entwickelt, das englische Princeton WordNet in RDF oder OWL zu konvertieren und damit für entsprechende Verfahren und Technologien des „Semantic Web“ nutzbar zu machen (Brickley(1999); Ciorăscu et al.(2003)Ciorăscu, Ciorăscu und Stoffel; van Assem et al.(2004)van Assem, Menken, Schreiber, Wielemaker und Wielinga; Graves und Gutierrez(2005)). Die *Web Ontology Language* OWL ist die W3C *recommendation* eines XML-basierten Standards zur Repräsentation von Ontologien im Semantic Web. Inzwischen liegt ein Public Working Draft einer W3C Arbeitsgruppe über ein standardisiertes Modell für die Repräsentation von Wortnetzen auf der Grundlage der genannten Arbeiten vor (van Assem et al.(2006b)van Assem, Gangemi und Schreiber).

In unserem Beitrag beschreiben wir, wie sich allgemeinsprachliche und domänenspezifische Wortnetze in OWL modellieren und miteinander verknüpfen lassen. Dementsprechend wurden repräsentative Teilausschnitte des in Tübingen entwickelten deutschen allgemeinsprachlichen Wortnetzes GermaNet und des in Dortmund entwickelten deutschen domänenspezifischen Wortnetzes TermNet modelliert und miteinander verlinkt. Ziel unseres Ansatzes ist es beide Netze vollständig in OWL zu konvertieren.

Unser Modellierungsvorschlag orientiert sich vor allem am Ziel der gemeinsamen Nutzung von Wortnetzen und Domänenontologien in texttechnologischen Anwendungen im Rahmen der DFG-Forschergruppe *Texttech-*

*nologische Informationsmodellierung.*

In Abschnitt 2 erörtern wir zunächst unseren Modellierungsansatz für das allgemeinsprachliche Wortnetz GermaNet, welcher weitgehend die W3C-Standardisierungsvorschläge für das Princeton WordNet aufgreift und diese an das Datenmodell von GermaNet adaptiert. Abschnitt 3 beschreibt die Modellierung für das terminologische Wortnetz TermNet, das semantische Bezüge zwischen Fachtermini aus den Domänen „Hypertextforschung“ und „Texttechnologie“ repräsentiert, die im Anwendungskontext der automatischen Hypertextualisierung genutzt werden. In Abschnitt 4 wird gezeigt, wie sich die beiden Ressourcen in Rahmen von OWL verknüpfen lassen. Wir orientieren uns dabei an dem sog. „Plug-in-Ansatz“ (Magnini und Speranza(2002)).

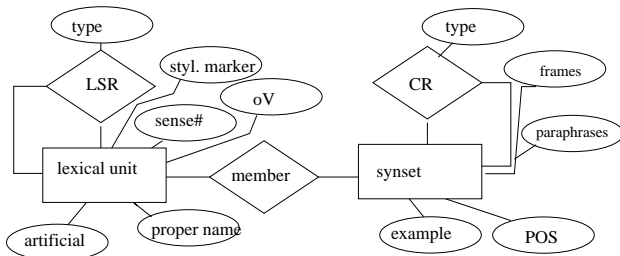
In Abschnitt 5 diskutieren wir eine Modellierungsalternative für GermaNet, in der Synsets und lexikalische Einheiten nicht als Instanzen (wie im W3C-Vorschlag), sondern als Klassen modelliert werden. Wir möchten zeigen, dass dies nicht nur die Nachbildung des Plug-in-Ansatzes im gegebenen Beschreibungsrahmen erleichtert, sondern auch semantische Intuitionen zum Verhältnis von Konzepten zu Instanzen adäquater widerspiegelt.

## 2 OWL-Modell für GermaNet

**GermaNet - Überblick und Datenstruktur** Mit GermaNet (Kunze(2001)) ist ein lexikalisch-semantisches Wortnetz im Stil des Princeton WordNet entwickelt worden, das die wichtigsten und häufigsten Konzepte des deutschen Grundwortschatzes abbildet und die grundlegenden semantischen Relationen zwischen Konzepten und lexikalischen Einheiten wie Hyperonymie, Meronymie und Antonymie modelliert. GermaNet enthält ca. 53 000 Synsets mit mehr als 76 000 Lexical Units. Einige Besonderheiten der GermaNet Modellierung betreffen:

- den Gebrauch nicht-lexikalischer sog. künstlicher Konzepte zur Bildung wohlgeformter Hierarchien;
- die Auszeichnung von Eigennamen, insb. aus dem Wortfeld geographischer Bezeichnungen;
- den taxonomischen Gliederungsansatz zur Modellierung der Adjektivhierarchien;
- die Modellierung der Meronymiebeziehung als generische Relation.

Das Datenmodell des GermaNet ist in Abb. 1 als ein Entity-Relationship-Graph wiedergegeben, welcher als Vorlage für eine XML-Modellierung des GermaNet (Lemnitzer und Kunze(2002)) diene.



CR=conceptual relation; LSR=lexical-semantic relation;oV=orthographic variant

Abbildung 1: E-R-Graph GermaNet

Im Diagramm sind die Objekte, Synsets und Lexical Units<sup>1</sup> als Rechtecke repräsentiert. Attribute, welche die lexikalischen Objekte charakterisieren, sind als Ellipsen gekennzeichnet. Die Relationen zwischen den Objekten werden als Rauten repräsentiert: konzeptuelle Relationen (CR) zwischen Synset Objekten wie z.B. Hyperonymie, und lexikalisch-semantic Relationen (LSR), die zwischen Lexical Units gelten, wie z.B. Antonymie.

**Klassenhierarchie** Aufgabe der OWL-Modellierung ist es, die Konzepte von GermaNet als Klassen (Konstrukt `<owl:class>`), sowie die Eigenschaften der Konzepte und die Relationen zwischen den Konzepten als Properties (Konstrukte `<owl:DatatypeProperty>` und `<owl:ObjectProperty>`) dieser Klassen darzustellen. Die grundlegenden Objekte des E-R-Modells führen unmittelbar zu den beiden Klassen *Synset* und *LexicalUnit* im OWL-Modell.<sup>2</sup>

Analog zum W3C-Vorschlag (van Assem et al.(2006a)van Assem, Gangemi und Schreiber) legen wir *NounSynset*, *VerbSynset*, *AdjectiveSynset* und *AdverbSynset* als unmittelbare Subklassen von *Synset* fest, sowie *NounUnit*, *VerbUnit*, *AdjectiveUnit* und *AdverbUnit* als unmittelbare Subklassen von *LexicalUnit*. In Ciorăscu et al.(2003)Ciorăscu, Ciorăscu und Stoffel) hat die der *LexicalUnit* entsprechende Klasse keine Subklassen, sie ergeben sich hier aber dadurch, dass es verschiedene CR sowie LSR gibt, die sich mit ihrem Definitionsbereich (*Domain*) und Wertebereich (*Range*) auf sie beziehen, etwa die Property *isParticipleOf* mit Domain *VerbUnit* und Range *AdjectiveUnit*. *NounSynset* hat eine weitere Subklasse *ProperName*, deren Instanzen Eigennamen sind.

**Properties** Die Verbindung zwischen Synsets und Lexical Units wird durch die Lexikalisierungsrelation hergestellt, wobei ein Synset durch mehrere Lexical Units lexikalisiert werden kann (die Synonyme). In OWL wird dies durch eine sog. *Object Property* namens *hasMember*, mit der Domain *Synset* und dem Range *LexicalUnit* modelliert. Für jede wortartbedingte Subklasse von *Synset* existiert eine Einschränkung des Wertebereichs von *hasMember* auf diejenige Subklasse von *LexicalUnit*, die der gleichen Wortart entspricht:

<sup>1</sup>Eine Lexical Unit ist nach Cruse(1986)) als ein Paar mit einer lexikalischen Form und einer einzigen Bedeutung charakterisiert. Ein Synset ist eine Menge mit mindestens einer Lexical Unit. Im selben Synset enthaltene Lexical Units sind synonym.

<sup>2</sup>OWL wird unterteilt in die drei unterschiedlich expressiven Subsprachen OWL Full, OWL DL und OWL Lite. Die meisten Systeme zur Verarbeitung von OWL-Ontologien beziehen sich auf OWL DL. Daher modellieren auch wir GermaNet zunächst im Rahmen der Subsprache OWL DL.

```
<owl:Class rdf:ID="NounSynset">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:allValuesFrom>
        <owl:Class rdf:about="#NounUnit"/>
      </owl:allValuesFrom>
    </owl:Restriction>
  </owl:Restriction>
  <owl:InverseFunctionalProperty
    rdf:ID="hasMember"/>
  </owl:InverseFunctionalProperty>
  </owl:Restriction>
</rdfs:subClassOf>
</owl:Class>
```

OWL eignet sich vorzüglich für eine Modellierung der beiden grundsätzlichen Relationstypen CR und LSR. Sowohl der Typ der konzeptuellen Relationen als auch der lexikalisch-semantic Relationen bestehen zwischen intern definierten Klassen, entsprechen somit *Object Properties* in OWL. Object Properties können wie Klassen hierarchisch angeordnet sein, dies wird durch das Konstrukt `<rdfs:subPropertyOf>` erreicht. Unsere Modellierung beinhaltet demnach zwei top-level Object Properties, *conceptualRelation* (Domain *Synset* und Range *Synset*) und *lexicalSemanticRelation* (Domain *LexicalUnit* und Range *LexicalUnit*). Die Charakterisierungen der jeweiligen Subproperties sind Tabelle 1 zu entnehmen.

Gleichgeordnet mit *conceptualRelation* und *lexicalSemanticRelation* befinden sich auf der oberen Ebene der Property-Hierarchie ebenfalls die bereits vorgestellten *hasMember* und *memberOf* sowie auch die Property *hasExample* zur Einführung von Beispielen für Synsets.

Auch die zentrale Relation Hyperonymie, die in lexikalisch-semantic Wortnetzen eine Hierarchie aufspannt, gehört zu den konzeptuellen Relationen. Hyperonymie wird als eine `<owl:TransitiveProperty>` mit der Domain *Synset* und dem Range *Synset* eingeführt.

```
<owl:TransitiveProperty rdf:about="#isHyperonymOf">
  <rdfs:subPropertyOf
    rdf:resource="#conceptualRelation"/>
  <rdfs:domain rdf:resource="#Synset"/>
  <rdfs:type rdf:resource=
    "http://www.w3.org/2002/07/owl#ObjectProperty"/>
  <owl:inverseOf>
    <owl:TransitiveProperty rdf:about="#isHyponymOf"/>
  </owl:inverseOf>
</owl:TransitiveProperty>
```

Gleichzeitig wird die ebenfalls existierende Hyponymie-Relation als zugehörige inverse Property deklariert. Ähnlich wie im Fall der Property *hasMember* gibt es für jede wortartbedingte Subklasse von *Synset* eine Einschränkung des Wertebereichs von *isHyponymOf* auf Synsets der gleichen Subklasse mittels des `<owl:allValuesFrom>`-Konstrukts.

Tabelle 1 bietet einen Überblick über alle für GermaNet relevanten Object Properties und ihrer in OWL modellierbaren Charakteristika und Einschränkungen. Gegenüber dem E-R-Modell in Abb. 1 konnten einige präzisere Spezifikationen vorgenommen werden. Die Bezeichner wurden im Wesentlichen beibehalten, aber mit Affixen wie *has*, *is* oder *of* versehen, um z.B. die Richtung einer Relation zu verdeutlichen.

Für die Properties *isHyperonymOf* und *hasMember* sind inverse Properties definiert. Diese bieten keine zusätzliche Information, können aber die Verarbeitung beschleunigen.

Relationen, die nicht zwischen intern definierten Klassen bestehen, sondern vielmehr einer intern definierten Klasse einen Wertebereich in Form eines XML-Schema-Datentyps wie *string* oder *boolean* zuordnen, werden als OWL Datatype Properties modelliert. Für GermaNet

Property	Domain	Range	Charakteristika	Inverse Property	Lokale Restriktionen
hasMember	Synset	LexicalUnit	invers-funktional	memberOf	wortartbezogen
memberOf	LexicalUnit	Synset	funktional	hasMember	wortartbezogen
<i>Konzeptuelle Relationen (CR)</i>					
isHyperonymOf	Synset	Synset	transitiv	isHyponymOf	wortartbezogen
isHyponymOf	Synset	Synset	transitiv	isHyperonymOf	wortartbezogen
isHolonymOf	NounSynset	NounSynset			
isMeronymOf	NounSynset	NounSynset			
isAssociatedWith	Synset	Synset			
entails	VerbSynset	VerbSynset			
causes	VerbSynset ⊔ AdjectiveSynset	VerbSynset			
<i>Lexikalisch-semantische Relationen (LSR)</i>					
hasAntonym	LexicalUnit	LexicalUnit	symmetrisch	hasAntonym	wortartbezogen
hasPertonym	LexicalUnit	LexicalUnit			
isParticipleOf	VerbUnit	AdjectiveUnit			
hasExample	Synset	Example			

Tabelle 1: Charakteristika der Object Properties

sind dies offensichtlich diejenigen Charakteristika von lexikalischen Einheiten und Synsets, die im E-R-Modell in Abb. 1 als Ellipsen dargestellt sind.<sup>3</sup>

Tabelle 2 bietet einen Überblick über Datatype Properties in der OWL-Modellierung von GermaNet mit Angabe von Domain, Range und Funktionseigenschaft.

### OWL-Kodierung eines repräsentativen Ausschnitts

Um OWL-Code für das vorgestellte Modell zu generieren und es auf Konsistenz zu testen, wurde ein repräsentativer Ausschnitt von 37 Synsets und 83 LexicalUnits aus GermaNet sowie Relationen zwischen ihnen als Ontologie in OWL kodiert,<sup>4</sup> in dem jede (Sub)Klasse und jede Property mindestens einmal instanziiert ist. Somit kann nun der entstandene OWL-Code als Spezifikation für die Konvertierungsroutinen verwendet werden, die erforderlich sind, um schließlich das gesamte GermaNet in OWL nach dem oben vorgestellten Modell zu konvertieren. Als Ausgangsformat für die Konvertierung bieten sich entweder die Prolog-Version oder die XML-Version von GermaNet (Lemmitzer und Kunze(2002)) an.

### 3 OWL-Modell für das terminologische Wortnetz TermNet

TermNet ist ein terminologisches Wortnetz, das mit dem Ziel entwickelt wurde, Ähnlichkeiten und Unterschiede zwischen Termini aus unterschiedlichen Terminologien zur selben Fachdomäne zu repräsentieren. Es wurde im Rahmen eines Projekts zur automatischen Hypertextualisierung aufgebaut und diente in der ersten Projektphase zur Generierung von Glossareinträgen in einem Fachtextkorpus (Beißwenger et al.(2003)Beißwenger, Storrer und Runte; Lenz et al.(2003)Lenz, Birkenhake und Maas). Das in der ersten Projektphase aufgebaute terminologische Netz enthält ca. 200 Termini aus den Fachdomänen „Texttechnologie“ und „Hypertextforschung“. Dabei wurden alle Termini aufgenommen, die im Korpus explizit definiert sind.

<sup>3</sup>Eine Ausnahme bildet *example*, da diese Eigenschaft aus zwei Komponenten besteht.

<sup>4</sup>Als Ontologie-Editor wurde Protégé 3.1.1 (<http://protege.stanford.edu>) in Verbindung mit der Reasoner/Classifier-Software RacerPro 1.9.1 (<http://www.racer-systems.com>) verwendet.

In der zweiten Phase des Projekts soll TermNet darüber hinaus für Strategien der automatischen Verlinkung nach thematischen Kriterien genutzt werden. Für derartige Verlinkungsstrategien, wie sie z.B. für das Englische auf der Basis von lexikalischen Ketten (*lexical chains*) vorgeschlagen wurden (Green(1999)), ist es notwendig, das terminologische Wortnetz mit einem großen allgemeinsprachlichen Wortnetz zu verbinden. Zu diesem Zweck wäre es wünschenswert, das terminologische Netz in OWL zu remodellieren und mit einer OWL-Repräsentation des GermaNet zu verknüpfen. Der folgende Modellierungsansatz wurde vor diesem Hintergrund entwickelt.

Bei der Übertragung des Wortnetzansatzes auf Fachwortschatz möchten wir folgenden Unterschieden zwischen Fach- und Allgemeinwortschatz Rechnung tragen:

1. In terminologischen Taxonomien können vertikale Relationen der Über- und Unterordnung meist als Klasseninklusion dargestellt werden. Wir nutzen deshalb die in OWL vordefinierte *subclassOf*-property, um die Unterordnung eines spezifischen Terminus, z.B. *MonodirektionaleVerknüpfung* unter einen generelleren Terminus, in diesem Fall *Verknüpfung*, zu repräsentieren.
2. Typisch für solche Taxonomien ist es, dass sich die einem generellen Terminus untergeordneten Termini nach dem relevanten klassifikatorischen Merkmal in wechselseitig disjunkte Subklassen gruppieren lassen. Zum Beispiel gibt es zum Terminus *Verknüpfung* die spezifischeren Termini *MonodirektionaleVerknüpfung*, *BidirektionaleVerknüpfung* (Merkmal: Direktionalität) und *IntratextuelleVerknüpfung*, *IntertextuelleVerknüpfung*, *ExtratextuelleVerknüpfung*. (Merkmal: Position des Linkziels). Die Termini, die nach demselben Merkmal klassifiziert sind, denotieren jeweils disjunkte Teilklassen. Diese Disjunktivität kann man, wenn man Termini als Klassen modelliert, durch entsprechende *<owl:disjointWith>*-Konstrukte festlegen.
3. Für viele Fachdomänen ist es charakteristisch, dass derselbe Untersuchungsbereich theorie- und schulenabhängig in unterschiedlicher Weise terminologisiert wird. Für den Terminus *Verknüpfung*, wie er im Standardwerk von Rainer Kuhlen (Kuhlen(1991)) definiert ist, finden sich in Dokumenten, die auf anderen Ansätzen basieren, entspre-

Property	Domain	Range	funktional
POS	Synset	„N“ „V“ „A“ „ADV“	ja
hasParaphrase	Synset	<xsd:string>	nein
isArtificial	Synset ⊆ LexicalUnit	<xsd:boolean>	(ja)
isProperName	NounSynset ⊆ NounUnit	<xsd:boolean>	(ja)
hasOrthographicForm	LexicalUnit	<xsd:string>	ja
hasSenseInt	LexicalUnit	<xsd:positiveInteger>	ja
isStylisticallyMarked	LexicalUnit	<xsd:boolean>	(ja)
hasFrame	VerbUnit ⊆ Example	<xsd:string>	nein
hasText	Example	<xsd:string>	ja

Tabelle 2: Charakteristika der Datatype Properties

chende Termini (z.B. *Link*, *Verweis*). Im Gegensatz zu lexikalischen Synonymen als Mitglieder desselben Synsets können formverschiedene Termini aus verschiedenen Taxonomien in den Fachtexten nicht wechselseitig ersetzt werden, auch wenn die denotierten Kategorien sehr ähnlich sind. Schließlich ist es ja gerade das Ziel von Terminologisierungprozessen, die Formseite des Terminus eindeutig an eine fachsprachliche Kategorie zu binden. Andererseits bestehen zwischen den Kategorien konkurrierender Systeme durchaus Entsprechungsbeziehungen. Z.B. haben die durch die Termini *Verknüpfung* (i.S. von Kühlen(1991))) und *Verweis* (i.S. von Tochtermann(1995))) denotierten Instanzen einen großen Schnittbereich. Um derartige Entsprechungen von Kategorien zu repräsentieren, nutzt TermNet, als fachsprachliche Entsprechung zu den Synsets der allgemeinsprachlichen Wortnetze, das Modellierungskonstrukt *TermSet*. Ein *TermSet* enthält die Termini einer Fachdomäne, die aus verschiedenen terminologischen Systemen stammen, aber eine sehr hohe Schnittmenge in den denotierten Instanzen haben.

Im OWL-Modell von TermNet werden TermSets als Klassen modelliert. Für die Darstellung hierarchischer Bezüge zwischen TermSets, also zwischen einem TermSet, das *Verknüpfung*, *Verweis* enthält, und einem spezifischeren TermSet, das *MonodirektionaleVerknüpfung*, *MonodirektionalerVerweis* enthält, möchten wir aber bewusst nicht den vordefinierten `subClassOf`-Konstruktor verwenden. Schließlich besteht die damit implizierte Klasseninklusion nur innerhalb desselben terminologischen Systems. Wir nutzen statt dessen hierfür die „weichere“ hierarchische Relation der Hyponymie und ihre Inverse Hyperonymie. Im Gegensatz zur Modellierung dieser Relation im OWL-Modell von GermaNet (s. Abschnitt 2) besteht die Hyponymie-Beziehung aber nicht zwischen Instanzen, sondern zwischen Klassen (TermSets). Sie muss also als Property definiert werden, die paarweise einzelne TermSets relationiert. Hierzu beschränken wir den Wertebereich unserer `tn:isHyponymOf`-Property durch das `<owl:allValuesFrom>`-Konstrukt, durch das die Menge aller Instanzen der betreffenden Klasse denotiert wird. Das folgende OWL-Statement drückt entsprechend aus, dass das TermSet *MonodirektionalerLink* ein Hyponym des TermSets *Link* ist.

```
<owl:Class rdf:ID="tn:TermSet_MonodirektionalerLink">
  <rdfs:subClassOf rdf:resource="#tn:NounTermSet" />
  ...
  <rdfs:subClassOf>
    <owl:Restriction>
```

```
<owl:onProperty rdf:resource="#tn:isHyponymOf" />
<owl:allValuesFrom
  rdf:resource="#tn:TermSet_Link" />
</owl:Restriction>
</rdfs:subClassOf>
...
</owl:Class>
```

Für die Zuordnung von Termini zu TermSets kann man nicht auf die `gn:hasMember`-Property zurückgreifen, wie sie im oben beschriebenen Modellierungsvorschlag für GermaNet definiert wurde. Denn dort sind ja – entsprechend dem Standardisierungsvorschlag für das Princeton WordNet – sowohl Lexical Units als auch Synsets als Instanzen modelliert. In TermNet hingegen ist jeder Terminus und jedes TermSet als Klasse definiert. Die TermNet-Entsprechung zur `gn:hasMember`-Property wird deshalb – ähnlich wie bei der Hyponymie – als eine paarweise Zuordnung von Klassen der Superklasse *NounTerm* (als Domäne) und Klassen der Superklasse *NounTermSet* (als Range) spezifiziert.

#### 4 Mögliche Schnittstellen zwischen Wortnetzen und Terminologien

Die Verknüpfung allgemeinsprachlicher mit domänenspezifischen Wortnetzen zur Ausnutzung der Vorteile beider Ressourcentypen in NLP und Semantic Web wird u.a. in den Ansätzen von Vossen(2001) und Magnini und Speranza(2002)) diskutiert. Während Vossen aus großen Dokumentensammlungen mit Verfahren der Informationserschließung und der Termextraktion ad-hoc Terminologien aufbaut, deren Termini anschließend an WordNet Synsets gelinkt werden, entwickeln Magnini und Speranza einen Ansatz zur Verknüpfung zweier bestehender Ressourcen, des ItalWordNet mit einer spezialisierten Ontologie aus der Domäne der Wirtschaft (ECOWN), den so genannten *Plug-in-Approach*. Verschiedene Plug-in-Relationen (Synonymy, Near-Synonymy, Hyponymy) dienen dazu, die jeweilig relevanten Synsets aus beiden Ressourcen zu vereinigen. Mittels relativ weniger Plug-ins können große Wortschatzmengen aus beiden Ontologien korreliert werden (269 Plug-in-Relationen für die Anbindung von 4662 ECOWN Konzepten an ItalWordNet). Das vorgestellte Verfahren führt zu einer Gesamthierarchie, welche die Topkonzepte der spezialisierten Ontologie ausblendet und ihre Unterkonzepte, die Termini, in die allgemeinsprachliche Ontologie integriert.

Der Plug-in Approach bietet eine attraktive Modellierungsvorlage für die Anbindung des TermNet an GermaNet, da beide Ressourcen ebenfalls wortnetzbasierend und von unterschiedlicher Größe bzw. Spezifität sind und sich in Bezug auf eine signifikante Anzahl von Konzepten überschneiden. Aufgrund der verschiedenartigen

Modellierung von Termini als Klassen und von Synsets als Instanzen in unserem Ansatz kann die Verknüpfung nur über die Einschränkung einer Property über die Klasse der Termini *NounTerm* als Domain erfolgen, da in OWL DL nicht die Klasse der Termini-Klassen als Metaklasse gebildet werden kann, d.h. Klassen nicht als Individuen betrachtet werden können. Wir unterscheiden die folgenden drei Fälle der Verlinkung von TermNet-Termini und GermaNet-Synsets:

1. Entsprechung von TermNet-Termini und GermaNet-Synsets: z.B. für den Terminus *tn:Term\_Link* das NounSynset *gn:Link*. Die Object Property *plg:attachedToNearSynonym* hat die Domain *tn:NounTerm* und den Range *gn:Synset*. Mittels einer `<owl:Restriction>` über *plg:attachedToNearSynonym* wird jedem Individuum der Klasse *tn:Term\_Link* das Individuum *tn:Link* zugeordnet.

```
<owl:Class rdf:ID="tn:Term_Link">
  <rdfs:subClassOf rdf:resource="#tn:NounTerm" />
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:hasValue rdf:resource="#gn:Link" />
      <owl:onProperty>
        <owl:ObjectProperty
          rdf:ID="plg:attachedToNearSynonym" />
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
  </owl:Class>
```

2. Ein TermNet-Terminus hat kein semantisches Synset-Pendant in GermaNet, aber der Terminus steht in einer Subclass-Relation zu einem Terminus, der über *plg:attachedToNearSynonym* an ein GermaNet-Synset angebunden ist. Z.B. ist der Terminus *tn:Term\_MonodirektionalerLink* über die Subclass-Property mit dem Terminus *tn:Term\_Link* verbunden, welcher wiederum über *plg:attachedToNearSynonym* an das NounSynset *gn:Link* in GermaNet geknüpft ist. Die Eigenschaft *plg:attachedToGeneralConcept* verbindet eine Terminusklasse wie *tn:Term\_MonodirektionalerLink* mit einem GermaNet-Synset, das mit einem übergeordneten Terminus in einer *plg:attachedToNearSynonym*-Relation steht. Die Relation zwischen indirekt verknüpften Konzepten kann somit auch direkt spezifiziert werden und dient der Verkürzung von Pfadlängen zwischen semantisch ähnlichen Konzepten, z.B. bei der Berechnung semantischer Distanzmaße. Hier gehen wir über den Plug-in Approach hinaus, der diese Relation nicht vorsieht.
3. Für einen Terminus aus TermNet gibt es kein semantisch entsprechendes Synset in GermaNet und auch kein passendes hyperonymisches GermaNet-Synset, an das der Terminus adäquat gelinkt werden könnte. Es existiert aber ein passendes holonymisches GermaNet-Synset, mit dem der TermNet-Terminus verknüpft werden kann, und zwar mittels der Plug-in-Relation *plg:attachedToHolonym*. Ein Beispiel ist der Terminus *tn:Term\_Anker*, der in seiner terminologischen Bedeutung in der Hypertextforschung (Anker = Teil eines Links) keine semantische Entsprechung in GermaNet hat, statt dessen aber über *plg:attachedToHolonym* dem meronymisch übergeordneten GermaNet-Synset *gn:Link* zugeordnet werden kann.

In den von uns bearbeiteten Ausschnitten konnten alle TermNet-Termini über eine der drei Plug-in-Properties an GermaNet angeschlossen werden. Bei der Anknüpfung weiterer Terminologien ist aber natürlich nicht auszuschließen, dass Termini weder eine Entsprechung noch einen passendes hyperonymisches oder holonymisches uebergeordnetes Synset in GermaNet haben. Nach einer möglichst vollständigen Verlinkung beider Ressourcen lässt sich ermitteln, wie häufig die unterschiedlichen Plug-in-Relationstypen jeweils instantiiert werden. Außerdem sollte sich abzeichnen, wieviele Konzepte mit den drei Plug-in-Relationen nicht anbindbar sind und ob sich die Einführung weiterer Relationstypen empfiehlt.

Durch die Modellierung der Termini als Klassen in TermNet und der Synsets als Instanzen in GermaNet sind die von uns vorgeschlagenen Schnittstellen-Relationen in OWL DL nicht aus der Sicht von GermaNet formulierbar, d.h. mit *Synset* als Domain. Dies ist für Inferenzen zwar nicht unbedingt erforderlich, aber dennoch wünschenswert.

## 5 Probleme und offene Fragen

Die von uns untersuchten OWL-Konvertierungen des Princeton WordNet (Ciorăscu et al.(2003)Ciorăscu, Ciorăscu und Stoffel; van Assem et al.(2004)van Assem, Menken, Schreiber, Wielemaker und Wielinga; van Assem et al.(2006a)van Assem, Gangemi und Schreiber) modellieren die einzelnen Synsets sowie die einzelnen Lexical Units als Instanzen. Aus linguistisch-lexikologischer Sicht ist dies verblüffend, da Synsets gemeinhin als Konzepte aufgefasst werden, die durch im Synset enthaltene Lexical Units lexikalisiert werden. Man hätte also erwartet, dass ein Synset für das Konzept /HUND/ als Klasse modelliert ist, der individuelle Hunde wie „Fido“ und „Waldi“ als Instanzen zugeordnet werden können, und dass konzeptuelle Relationen zwischen solchen Klassen definiert werden. Die Entscheidung, Synsets und Lexical Units als Instanzen zu modellieren, scheint verarbeitungstechnisch und nicht linguistisch motiviert zu sein: So weist z.B. Sanfilippo et al.(2005)Sanfilippo, Tratz, Gregory, Chappell, Whitney, Posse, Paulson, Baddeley, Hohimer und White) darauf hin, die Modellierung einer großen Zahl von Lexical Units und Synsets als Klassen sei „impractical for a real-world application“. Wir haben uns mit dem in Abschnitt 2 vorgestellten OWL-Modell für GermaNet dieser Entscheidung angeschlossen, möchten aber, weil diese aus linguistischer Perspektive doch recht unbefriedigend ist, weiter mit anderen Modellierungsoptionen experimentieren.

Allerdings erfordert die Definition einer OWL Property, die eine Menge von Klassen paarweise mit einer anderen Menge von Klassen relationiert (wie bei der Meronymie/Holonymie und anderen konzeptuellen Relationen), die Verwendung von Meta-Klassen, wie sie nur in OWL Full zur Verfügung stehen. Wenn man im Rahmen von OWL DL bleiben möchte, gibt es noch die Möglichkeit, Subklassen durch Einschränkung einer Relation von einer Superklasse auf diese Subklassen einander paarweise zuzuordnen, wie am Beispiel der Hyperonymie-Relation in TermNet gezeigt.

Unklar ist, wie sich eine solche Modellierung für große Datenmengen verarbeiten lässt. Wir haben vor, dies an einem größeren Ausschnitt von GermaNet

zu testen. Eine solche alternative Modellierung hätte auch den Vorteil, dass die Relation *hyperonymy* mit `<rdfs:subClassOf>` gleichgesetzt werden kann, was für bestimmte Anwendungen interessant sein kann, vgl. (van Assem et al.(2006b)van Assem, Gangemi und Schreiber). Außerdem könnte bei einer solchen Modellierung der Plug-in-Ansatz zur Verknüpfung von allgemeinsprachlichen und terminologischen Wortnetzen eleganter umgesetzt werden, weil dann zu den Properties, die beide Ressourcen verknüpfen, auch innerhalb von OWL DL jeweils inverse Properties angelegt werden können. Sollte eine Modellierung aller Synsets und Lexical Units als Klassen nicht machbar sein, bietet sich eine Teillösung an: Interessant in diesem Zusammenhang ist z.B. der Vorschlag von Sanfilippo et al.(2005)Sanfilippo, Tratz, Gregory, Chappell, Whitney, Posse, Paulson, Baddeley, Hohimer und White), nur eine geringere Teilmenge (1077) der Synsets, und zwar die generischen Topkonzepte, als OWL-Klassen zu modellieren und die restlichen Synsets auf diese abzubilden.

Neben der Entwicklung und Evaluierung linguistisch adäquaterer Modellierungsalternativen werden wir den Ansatz zur Verknüpfung von GermaNet mit terminologischen Wortnetzen detaillierter ausarbeiten und den Ansatz auch für andere Domänenontologien austesten. Die entstehenden Ressourcen werden in texttechnologischen Anwendungen evaluiert.

## References

Beißwenger, Michael; Storrer, Angelika und Runte, Maren (2003): "Modellierung eines Terminologienetzes für das automatische Linking auf der Grundlage von WordNet". In: *Anwendungen des deutschen Wortnetzes in Theorie und Praxis*. Tübingen, S. 113–126.

Brickley, Dan (1999): "Wordnet in RDF/XML: 50,000+ RDF class vocabulary". Message to RDF Interest Group. <http://lists.w3.org/Archives/Public/www-rdf-interest/1999Dec/002.html>.

Ciorăscu, Iulian; Ciorăscu, Claudia und Stoffel, Kilian (2003): "Scalable ontology implementation based on knOWLer". In: *Proceedings of the 2nd International Semantic Web Conference (ISWC2003), Workshop on Practical and Scalable Semantic Systems*. Sanibel Island, Florida.

Cruse, D. Alan (1986): *Lexical Semantics*. Cambridge: Cambridge University Press.

Fellbaum, Christiane (Herausgeber) (1998): *WordNet – An Electronic Lexical Database*. Language, Speech, and Communication. Cambridge, MA: MIT Press.

Graves, Alvaro und Gutierrez, Claudio (2005): "Data Representation for WordNet: A case for RDF". In: *Proceedings of the Third International WordNet conference, Jeju Island, Korea*.

Green, Stephen J. (1999): "Lexical semantics and automatic hypertext construction". *ACM Computing Surveys* 31 (4).

Kuhlen, Rainer (1991): *Hypertext. Ein nicht-lineares Medium zwischen Buch und Wissensbank*. Berlin.

Kunze, Claudia (2001): "Lexikalisch-semantische Wortnetze". In: *Computerlinguistik und Sprachtechnologie: eine Einführung*, herausgegeben von Carstensen, Kai-Uwe et al., Heidelberg: Spektrum Verlag, S. 386–393.

Kunze, Claudia; Lemnitzer, Lothar und Wagner, Andreas (2003): *Anwendungen des deutschen Wortnetzes in Theorie und Praxis. Sonderheft der Zeitschrift für Computerlinguistik und Sprachtechnologie*. Bonn.

Lemnitzer, Lothar und Kunze, Claudia (2002): "Adapting GermaNet for the Web". In: *Proceedings of the first Global WordNet Conference, Central Institute of Indian Languages, Mysore, India*. S. 174–181.

Lenz, Eva Anna; Birkenhake, Benjamin und Maas, Jan Frederik (2003): "Von der Erstellung bis zur Nutzung: Wortnetze als XML Topic Maps". In: *Anwendungen des deutschen Wortnetzes in Theorie und Praxis*. Tübingen, S. 127–136.

Magnini, Bernardo und Speranza, Manuela (2002): "Merging Global and Specialized Linguistic Ontologies". In: *Proceedings of Ontolex 2002. Las Palmas de Gran Canaria, Spain*. S. 43–48.

Sanfilippo, Antonio; Tratz, Stephen; Gregory, Michelle; Chappell, Alan; Whitney, Paul; Posse, Christian; Paulson, Patrick; Baddeley, Bob; Hohimer, Ryan und White, Amanda (2005): "Automating ontological annotation with wordnet". In: *Proceedings of the 5th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot2005) located at the 4th Semantic Web Conference*. Galway, Ireland.

Tochtermann, Klaus (1995): *Ein Modell für Hypermedia: Beschreibung und integrierte Formalisierung wesentlicher Hypermediakonzepte*. Aachen.

van Assem, Mark; Gangemi, Aldo und Schreiber, Guus (2006a): "Conversion of WordNet to a standard RDF/OWL representation". In: *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. Genoa, Italy. In press.

van Assem, Mark; Gangemi, Aldo und Schreiber, Guus (2006b): "RDF/OWL Representation of WordNet". W3C Public Working Draft of 19 June 2006 of the Semantic Web Best Practices and Deployment Working Group. <http://www.w3.org/TR/wordnet-rdf/>.

van Assem, Mark; Menken, Maarten R.; Schreiber, Guus; Wielemaker, Jan und Wielinga, Bob (2004): "A method for converting thesauri to RDF/OWL". In: *Proceedings of the 3rd International Semantic Web Conference (ISWC 2004)*. Hiroshima, Japan, Nummer 3298 in Lecture Notes in Computer Science.

Vossen, P. (2001): "Extending, trimming and fusing WordNet for technical documents". In: *Proceedings of NAACL-2001 Workshop on WordNet and Other Lexical Resources Applications*. Pittsburgh, USA.