

**ISPRI-Arbeitsbericht  
Nr. 01/2006**

**Ontologie-basierte Attributierung  
von Informationsmodellen:**

**Grundlagen und Anwendungsgebiete**

**Ahlemann, F.; Teuteberg, F.; Brune, G.**

**Herausgeber:**

Frank Teuteberg; Frederik Ahlemann

## **Die Herausgeber**

Jun.-Prof. Dr. Frank Teuteberg  
Universität Osnabrück  
Lehrstuhl für BWL/E-Business und Wirtschaftsinformatik  
D-49069 Osnabrück

Dr. Frederik Ahlemann  
Universität Osnabrück  
Lehrstuhl für BWL/Organisation und Wirtschaftsinformatik  
D-49069 Osnabrück

## **ISPRI-Arbeitsbericht Nr. 01/2006**

© ISPRI – Forschungszentrum für Informationssysteme in Projekt- und Innovationsnetzwerken  
<http://www.ispri.net>

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der Grenzen des Urhebergesetzes ist ohne Zustimmung der Herausgeber unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.  
Alle Rechte vorbehalten.

Osnabrück, im August 2006

ISBN 3-936475-26-1

## Die Autoren



Herr Dr. **Frederik Ahlemann**

E-Mail: frederik.ahlemann@ispri.net

Dr. Frederik Ahlemann studierte Wirtschaftsinformatik an der Universität Münster und war danach als Berater im Bereich Projektmanagement beschäftigt. Dann wechselte er mit dem Ziel der Promotion zur Universität Osnabrück an den Lehrstuhl für Betriebswirtschaftslehre, Organisation und Wirtschaftsinformatik. Dort beschäftigt sich Herr Ahlemann mit Multi-Projektmanagement und der Konzeption entsprechender Anwendungssysteme. Er ist Autor einer umfassenden Marktstudie zu Projektmanagement-Softwaresystemen und federführend bei der Entwicklung einer DIN-Norm für den Austausch von Projektmanagementdaten. Darüber hinaus berät er Unternehmen bzgl. der Gestaltung von Projektmanagementprozessen und der Auswahl entsprechender Software. Im Jahr 2005 promovierte er zum Thema Referenzmodelle für das Projektmanagement.



Herr Jun.-Prof. Dr. **Frank Teuteberg**

E-Mail: frank.teuteberg@ispri.net

Frank Teuteberg studierte Betriebswirtschaftslehre mit Schwerpunkt Wirtschaftsinformatik an der Universität Göttingen. Nach seinem Abschluss im April 1996 war er von 1996 bis 2001 als wissenschaftlicher Mitarbeiter und nach seiner Promotion zum Dr. rer. pol. im Mai 2001 als wissenschaftlicher Assistent am Lehrstuhl für Wirtschaftsinformatik an der Europa-Universität Frankfurt (Oder) tätig. Seit April 2004 ist Herr Teuteberg Junior-Professor an der Juniorprofessur BWL/E-Business und Wirtschaftsinformatik der Universität Osnabrück. Frank Teuteberg ist Verfasser von 55 wissenschaftlichen Publikationen in z.T. führenden deutschen und internationalen Fachzeitschriften, darunter die Zeitschrift für Wirtschaftsinformatik, Electronic Markets: The International Journal of Electronic Commerce & Business Media und das International Journal of Computer Systems Science & Engineering. Seine Forschungsschwerpunkte sind Projekt- und Innovationsnetzwerke, Mobile Supply Chain Management, E-Negotiations und Market Engineering sowie Semantic Mining von Informationsmodellen.



Herr Diplom-Kaufmann **Guido Brune**

E-Mail: gbrune@uos.de

Guido Brune studierte Betriebswirtschaftslehre mit Schwerpunkt Wirtschaftsinformatik an der Universität Osnabrück. Herr Brune arbeitet derzeit als Controller bei der DORMA-Glas GmbH.

## **Zusammenfassung**

In dem vorliegenden Beitrag wird gezeigt, wie durch eine zielgerichtete ontologie-basierte Attributierung von Informationsmodellen die Grundlage für eine automatisierte bzw. halbautomatisierte Konstruktion, Interpretation und Verarbeitung von Informationsmodellen auf semantischer Ebene geschaffen werden kann. Aus dem dargestellten Konzept ergibt sich eine Reihe von innovativen Anwendungsbereichen, die bei einer herkömmlichen subjektgebundenen Interpretation von Informationsmodellen nicht realisierbar sind. Entsprechende Ansätze wie „Information Model Mining“, „Information Model Based Benchmarking“ oder „Semi-Automatic Reference Model Construction“ werden skizziert. Weiterer Forschungsbedarf wird aufgezeigt.

## **Schlüsselwörter**

Informationsmodelle, Meta-Modellierung, Distanz- und Ähnlichkeitsmaße, Metriken, Ontologien, Referenz-Metamodell, Benchmarking, Referenzmodellierung

## **Vorwort**

Dies ist der zweite Arbeitsbericht des Forschungszentrums für Informationssysteme in Projekt- und Innovationsnetzwerken (ISPRI) an der Universität Osnabrück.

Der vorliegende Arbeitsbericht setzt eine Berichtsreihe fort, welche die Forschungsergebnisse des ISPRI dokumentiert.

Herausgeber der Reihe sind die wissenschaftlichen Sprecher des ISPRI Jun.-Prof. Dr. Frank Teuteberg und Dr. Frederik Ahlemann.

Das Ziel dieses Arbeitsberichts ist es, die theoretischen Grundzüge einer ontologie-basierten Attributierung von Informationsmodellen zu erarbeiten, welche die Grundlage für eine automatisierte bzw. halbautomatisierte Konstruktion, Interpretation und Verarbeitung von Informationsmodellen auf semantischer Ebene schafft und damit neue Einsatzfelder von Informationsmodellen ermöglicht.

Frederik Ahlemann, Frank Teuteberg

Osnabrück, im August 2006

# **Ontologie-basierte Attributierung von Informationsmodellen: Grundlagen und Anwendungsgebiete**

*Frederik Ahlemann, Frank Teuteberg, Guido Brune*

# Inhaltsverzeichnis

<b>1</b>	<b>EINLEITUNG</b>	<b>1</b>
<b>2</b>	<b>KONZEPTIONELLE GRUNDLAGEN</b>	<b>3</b>
2.1	Informationsmodelle	3
2.2	Metamodelle und Meta-Metamodelle	3
2.3	Ontologien	5
<b>3</b>	<b>ÜBERBLICK</b>	<b>8</b>
<b>4</b>	<b>ENTWICKLUNG EINES SYNTAKTISCHEN META-METAMODELLS</b>	<b>9</b>
<b>5</b>	<b>ENTWICKLUNG EINER PROBLEMANGEMESSENEN DOMÄNENONTOLOGIE</b>	<b>12</b>
<b>6</b>	<b>SEMANTISCHE ATTRIBUTIERUNG VON INFORMATIONSMODELLEN</b>	<b>14</b>
<b>7</b>	<b>AUTOMATISIERTER VERGLEICH VON INFORMATIONSMODELLEN</b>	<b>16</b>
7.1	Metriken	16
7.2	Distanz- und Ähnlichkeitsmaße	21
7.3	Normierung und Skalierung	22
<b>8</b>	<b>ARCHITEKTUR EINES ALLGEMEINEN INFORMATIONSMODELL-DATENBANKSYSTEMS</b>	<b>23</b>
<b>9</b>	<b>ANWENDUNGSBEREICHE</b>	<b>26</b>
9.1	Informationsmodell-basiertes Benchmarking	26
9.2	Erfahrungsdatenbanken	27
9.3	Information Model Mining	28
9.4	(Halb-)Automatische Referenzmodellkonstruktion	28
9.5	Automatisierte Softwareauswahl	29
<b>10</b>	<b>ZUSAMMENFASSUNG UND AUSBLICK</b>	<b>29</b>
	<b>LITERATUR</b>	<b>31</b>

## Abbildungsverzeichnis

Abb. 1: Metamodelle und Meta-Metamodelle	4
Abb. 2: Informationsmodell	6
Abb. 3: Beispiel zu Meta-Metamodellen	11
Abb. 4: Beispiel zur Ontologie-Konstruktion	13
Abb. 5: Konflikttypen beim Vergleich von Informationsmodellen	15
Abb. 6: Semantische Attributierung auf Basis von Ontologien	15
Abb. 7: Metrik zum Informationsmodellvergleich	20
Abb. 8: Ontologie-basierte Attributierung als UML-Diagramm	24
Abb. 9: Grundzüge der Architektur eines Informationsmodell-Datenbank-systems	25

## Tabellenverzeichnis

Tab. 1: Ausgewählte grundlegende Metriken zum Vergleich von Ontologien	18
Tab. 2: Kohäsions- und Komplexitätsmetriken zum Vergleich von Ontologien	19
Tab. 3: Distanzmaße zum Vergleich von Informationsmodellen mittels Merkmalsvektoren	22

# 1 Einleitung

Informationsmodelle haben sich als Grundlage für Vorhaben der Informationssystem- und Organisationssystemgestaltung durchgesetzt. Darüber hinaus werden sie zunehmend auch im Rahmen des Wissensmanagement eingesetzt. Die Erstellung und Anwendung von Informationsmodellen erfordert die Involvierung eines Subjektes, das mit seinen Kenntnissen und Fähigkeiten die Konstruktion und die Interpretation des Informationsmodells erst möglich macht. Die einzelnen Modellbestandteile werden dabei jeweils mit den internen Modellen des Subjektes vernetzt, so dass eine Zuordnung zu Bekanntem und damit eine handlungsorientierte Verarbeitung der im Modell kodierten Informationen erfolgen kann. Aufgrund dieser Notwendigkeit der Vernetzung mit den subjektgebundenen internen Modellen ist die Modellkonstruktion und -interpretation nicht beliebig skalierbar; jede wiederholte Nutzung eines Informationsmodells erfordert eine wiederholte subjektgebundene Interpretation seiner Inhalte. Das bedeutet, dass jede wiederholte Nutzung eine Person oder ggf. sogar einen Domänenexperten erfordert. Für eine Reihe von innovativen Informationsmodellierungen ist jedoch eine automatisierte bzw. halbautomatisierte Konstruktion bzw. Interpretation von Modellen wünschenswert oder erforderlich. So lassen sich durch die Möglichkeit einer automatisierten Verarbeitung Anwendungsbereiche wie *Information Model Mining*, *Information Model Based Benchmarking* oder *Semi-Automatic Reference Model Construction* erschließen.<sup>1</sup> Hierzu ist es erforderlich, zusätzlich zur syntaktischen Analyse auch die semantische Analyse von Informationsmodellen zu ermöglichen. Während Erstere durch Metamodelle seit langem möglich ist und praktiziert wird, erfordert Letztere zusätzliche Informationen über den Modellgegenstand bzw. den Realweltausschnitt, der im Modell (re-)konstruiert wird. Hierbei können Ontologien hilfreich sein. Durch eine Zuordnung von Ontologiekonzepten zu den Elementen des Informationsmodells ist die Grundlage für eine semantische Analyse geschaffen. Verwandte konzeptionelle Ansätze existieren im Bereich des *Enterprise Application Integration* (EAI) bzw. *Information Integration*. Hier werden unter dem Begriff *Ontology Merging*<sup>2</sup> Ansätze bezeichnet, bei denen auf Basis einer übergeordneten Ontologie, die aus divergierenden Ontologien konstruiert wird, die semantische Integration von Datenstrukturbeschreibungen bzw. Datenbankschemata erreicht wird.<sup>3</sup> Auf diese Weise wird eine inner- bzw. zwischenbetriebliche Integration von heterogenen Informationssystemen ermöglicht.

---

<sup>1</sup> Eine ausführliche Diskussion dieser Anwendungsbereiche erfolgt in Abschnitt 9.

<sup>2</sup> Vgl. Noy, N.F., Musen, M.A. (1999).

<sup>3</sup> Vgl. beispielhaft Bergamaschi, S. et al. (2001); Preece, A.D. et al. (2001), S. 171-195; ein ausführlicher Vergleich unterschiedlicher Ansätze erfolgt bei De Bruijn, J. et al. (2004).

Das Ziel dieser Arbeit ist es, die theoretischen Grundzüge einer ontologie-basierten Attributierung von Informationsmodellen zu erarbeiten, welche die Grundlage für eine automatisierte Konstruktion, Interpretation und Verarbeitung von Informationsmodellen auf semantischer Ebene schafft und damit neue Einsatzfelder von Informationsmodellen ermöglicht.

Die Arbeit ist wie folgt gegliedert: In Abschnitt 2 werden begriffliche und konzeptionelle Grundlagen erörtert. Darüber hinaus werden wichtige Vorarbeiten präsentiert. Abschnitt 3 gibt einen Überblick über das Konzept der semantischen Attributierung von Informationsmodellen und beschreibt den zugrunde liegenden Verarbeitungsprozess. Die folgenden Abschnitte 4 bis 7 behandeln in detaillierter Weise Einzelaspekte des Konzeptes. Darüber hinaus wird eine Präzisierung in Form einer mathematischen Formalisierung vorgenommen. Danach wird im Abschnitt 8 die Architektur einer allgemeinen Informationsmodell-Datenbank beschrieben, die das hier skizzierte Konzept technisch unterstützt. In Abschnitt 9 werden mögliche Anwendungsbereiche des Konzeptes skizziert. Abschnitt 10 fasst die Ergebnisse zusammen und zeigt weitere Forschungsbedarfe auf.

## 2 Konzeptionelle Grundlagen

### 2.1 Informationsmodelle

Informationssysteme sind Systeme, deren primärer Zweck die Informationsbereitstellung ist. Sie leisten eine Vielzahl von Einzelaufgaben, wie z.B. Erfassung, Bearbeitung, Verarbeitung, Speicherung, Transport und Ausgabe von Informationen.<sup>4</sup> Unter einem Informationsmodell wird das Ergebnis einer Konstruktion eines Modellierers verstanden, der für Anwendungssystem- und Organisationsgestalter Informationen über zu modellierende Elemente eines Informationssystems zu einer Zeit mit Hilfe einer Sprache als relevant deklariert.<sup>5</sup>

Informationsmodelle können in verschiedene Klassen eingeteilt werden:<sup>6</sup> Aus der Perspektive der Systemtheorie können in einem Informationsmodell zwei verschiedene *Sichten* eingenommen werden. Einerseits kann untersucht werden, welche Informationsstrukturen vorliegen (Struktursicht), andererseits kann die Erzeugung, Verarbeitung und Nutzung von Informationen Gegenstand der Betrachtung sein (Verhaltenssicht). Beide Sichten münden in der gängigen Unterscheidung von Daten und Prozessen.

Eine weitere Einteilung ergibt sich, wenn Modelle nach ihrer *inhaltlichen Breite* unterschieden werden. Anwendungsinformationsmodelle beschreiben die Situation eines spezifischen Unternehmens zu einer Zeit, während Referenzinformationsmodelle Sollcharakter für eine Klasse von Unternehmen besitzen und einer unternehmensspezifischen Anpassung bedürfen.

### 2.2 Metamodelle und Meta-Metamodelle

Wird der *Abstraktionsgrad* eines Modells zugrunde gelegt, kann zwischen Objektmodellen, Metamodellen<sup>7</sup> und Meta-Metamodellen unterschieden werden. Diese Klassifikation ergibt sich aus der Aussagestufe der Modellierungssprache. Diesbezüglich wird von Objekt- und Metasprachen gesprochen. Metasprachen<sup>8</sup> sind Sprachen, die über ei-

---

<sup>4</sup> Vgl. Heinrich, L.J. (2001), S. 14.

<sup>5</sup> Diese Definition basiert auf jener von Schütte, R. (1998), S. 63. Im Gegensatz zur Definition Schüttes wird hier explizit vom Informationssystem als „Modellgegenstand“ gesprochen; Schütte spricht von „System“.

<sup>6</sup> Vgl. Schütte, R. (1998), S. 63 ff.

<sup>7</sup> Der Begriff Metamodell hat sich durch prominente Metamodellierungsarchitekturen wie bspw. MOF (*Meta Object Facility*) durchgesetzt; vgl. Object Management Group (2003), <http://www.metamodel.com> sowie Meta Data Coalition unter <http://www.mdcinfo.com>.

<sup>8</sup> Die Silbe "*Meta*" stammt aus dem Altgriechischen und ist ein Präfix, das "darüber hinaus" bedeutet; vgl. Strahinger, R. (1998).

ner Objektsprache stehen und diese zum Gegenstand haben. Auf der Ebene der Modelle bedeutet dies, dass in Metamodellen Eigenschaften von Objektmodellen konstruiert werden. Werden Metamodelle zum Gegenstand einer Modellkonstruktion, wird von einem Meta-Metamodell gesprochen. Höhere Aussagestufen können mit dem Begriff  $Meta^n$ -Modell bezeichnet werden.<sup>9</sup>

Die *Richtigkeit* der Anwendung einer Darstellungstechnik und -sprache ergibt sich aus der fehlerfreien Einhaltung ihrer syntaktischen Regeln, die durch die einem Metamodell inhärente formale Grammatik beschrieben wird. Richtig ist ein Modell genau dann, wenn es vollständig und konsistent zum Metamodell ist. Von *Vollständigkeit* wird gesprochen, wenn die vom Metamodell geforderten Informationsobjekttypen und Beziehungen im Modell vollständig enthalten sind. *Konsistent* ist ein Modell, wenn seine Informationsobjekte vollständig im Metamodell spezifiziert sind.<sup>10</sup>

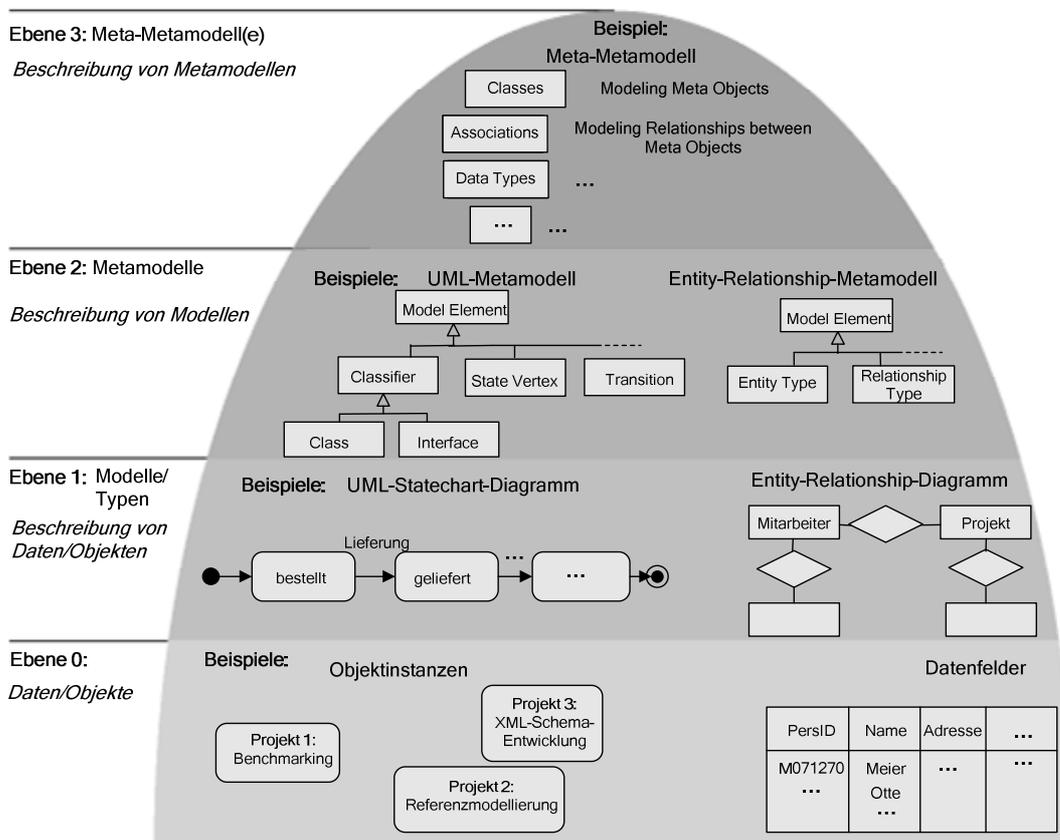


Abb. 1: Metamodelle und Meta-Metamodelle<sup>11</sup>

Sowohl die Ausschnitte der Welt, die durch die Informationsmodelle beschrieben werden als auch die Modelle selbst sind i.d.R. sehr heterogen. Um diese Heterogenität zu

<sup>9</sup> Vgl. Lehner, F. (1995), S. 39.

<sup>10</sup> Vgl. Schütte, R. (1998), S. 126.

<sup>11</sup> In Anlehnung an Fraunhofer ISST Jahresbericht 2003, S. 17.

überwinden, notwendige Modelltransformationen durchführen zu können sowie die modellübergreifende Konsistenz, die Richtigkeit und die Vollständigkeit von Modellen zu gewährleisten, ist es notwendig auf einer Metaebene oder einer Meta<sup>n</sup>-Ebene zu operieren und tiefere Ebenen der Modellierung formal zu beschreiben (vgl. Abbildung 1). Augenfällig ist, dass sowohl in der einschlägigen Literatur<sup>12</sup> als auch in der Praxis durch die UML und verwandte OO-Modellierungssprachen neuerer Generation im Wesentlichen die *Syntax* der Modellierungssprachen im Mittelpunkt der Betrachtung metamodellbasierter Ansätze (wie z.B. MOF,<sup>13</sup> UML-Metamodell,<sup>14</sup> CDIF-Meta-Metamodell<sup>15</sup>) stehen. Für eine automatische bzw. halbautomatische Interpretation und Verarbeitung von Informationsmodellen sind jedoch Metaisierungsprinzipien wünschenswert bzw. notwendig, welche die *Semantik* stärker in das Zentrum der Betrachtung rücken. Genau hier soll die in diesem Arbeitsbericht diskutierte ontologie-basierte Attributierung von Informationsmodellen ansetzen.

## 2.3 Ontologien

Der Begriff Ontologie wird in mehreren Wissenschaftsdisziplinen mit unterschiedlicher Bedeutung verwendet. In der Philosophie gehen die Wurzeln des Begriffs auf Aristoteles (384-322 v. Chr.) zurück. Entsprechend der griechischen Bedeutung von "*ontos*" (Sein, Ding, Objekt) und "*logos*" (Wort) ist die klassische Ontologie die "*Lehre vom Sein*" und seit Aristoteles die Grunddisziplin der Metaphysik.<sup>16</sup> Der Begriff Ontologie wird erstmals 1613 von den Philosophen Rudolf Göckel (im *Lexicon philosophicum*) und Jacob Lorhard (im *Theatrum philosophicum*) verwendet.<sup>17</sup>

Die wohl am häufigsten zitierte Definition des Begriffs Ontologie ist jene von Gruber, der eine Ontologie als "*an explicit specification of a conceptualization*"<sup>18</sup> beschreibt.

Ontologien sind:<sup>19</sup>

- *formal*, d.h. maschinenlesbare Modelle, die durch eine Menge von Klassen (Konzepten), Relationen, Instanzen und logischen Aussagen beschrieben werden,
- *explizit*, d.h. die verwendeten Konzepte, deren Randbedingungen sowie Datentypen werden explizit spezifiziert,

---

<sup>12</sup> Vgl. bspw. Object Management Group (1999); Object Management Group (2003).

<sup>13</sup> Vgl. Object Management Group (2003).

<sup>14</sup> Vgl. Object Management Group (1999).

<sup>15</sup> Vgl. Electronic Industries Association (1994).

<sup>16</sup> Aristoteles postulierte eine Wissenschaft, die das "Seiende" untersucht und stellte diese als die Grundlegendste aller Wissenschaften dar.

<sup>17</sup> Vgl. Scherer, G. (1996), S. 16.

<sup>18</sup> Vgl. Gruber, T. R. (1993).

<sup>19</sup> Vgl. Gruber, T. R. (1993).

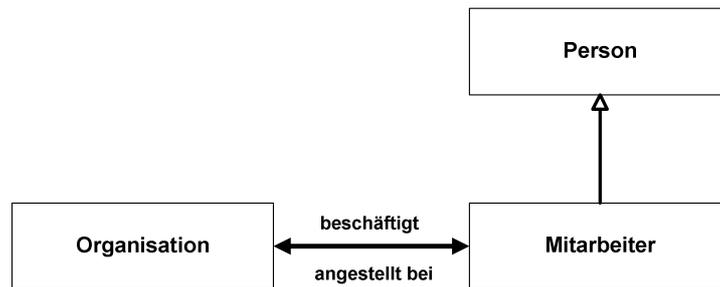
- *allgemein*, d.h. der Geltungsbereich einer Ontologie wird von einer Gruppe von menschlichen und maschinellen Akteuren als Kommunikationsgrundlage akzeptiert.

Eine formale Beschreibung einer Ontologie ist die folgende:<sup>20</sup>

*Eine Ontologie  $O$  ist ein 7-Tupel  $\{K, R, I, A, T, H^K, rel\}$ , wobei  $K$  eine Menge von Konzepten,  $R$  eine Menge von Relationen,  $I$  eine Menge von Instanzen,  $A$  eine Menge von logischen Aussagen (Axiome) über die verwendeten Objekte in einer wählbaren Logiksprache,  $T$  eine Menge von Termen (Vokabular),  $H^K \subseteq K \otimes K$  eine Konzepthierarchie, wobei  $H^K(K_1, K_2)$  bedeutet, dass  $K_1$  eine Unterklasse von  $K_2$  ist, sowie  $rel: R \rightarrow K \otimes K$  eine Funktion ist, die zwei Konzepte miteinander assoziiert. Die Mengen  $K, R, I, A, H^K$  sind disjunkt.*

Allgemeinere Konzepte werden in der Hierarchie höher angeordnet. Die einzelnen Konzepte können über Konstrukte wie "is-a"-Relationen, "part-of"-Relationen, "instance-of"-Relationen, subClassOf-Relationen, Kompositionen oder Aggregationen miteinander assoziiert werden.

Anhand eines Auszugs aus einem Informationsmodell (vgl. Abbildung 2) als Beispiel soll die formale Beschreibung verdeutlicht werden.



**Abb. 2: Informationsmodell**

Auf Basis unserer formalen Beschreibung einer Ontologie ergeben sich somit bspw. die folgenden Mengen und Funktionen:

- $K = \{Person, Organisation, Mitarbeiter\}$ ,
- $R = \{angestellt\_bei, beschäftigt, subClassOf\}$ , wobei "beschäftigt" eine inverse Relation zu "angestellt bei" ist.
- $H^K = (Mitarbeiter, Person)$ ,
- $I = (Frederik Ahlemann, Frank Teuteberg, Guido Brune, Forschungszentrum ISPRI, Juniorprofessur BWL/E-Business und Wirtschaftsinformatik,...)$
- $T = \{angestellt\_bei, beschäftigt, Person, \dots\}$
- $rel(angestellt\_bei) \rightarrow K_3 \otimes K_2$

<sup>20</sup> Vgl. Ehrig, M. et al. (2005), S. 508 f.

- $rel(beschäftigt) \rightarrow K_2 \otimes K_3$
- $rel(subClassOf) \rightarrow K_3 \otimes K_1$
- $A = \{ "FORALL X, Y : Mitarbeiter: X [angestellt\_bei \rightarrow Y] AND Organisation: Y [beschäftigt \rightarrow X]", \dots \}$ .

Mittels einer Ontologie wird somit eine bestimmte Sichtweise auf Entitäten (wie z.B. Daten, Objekte, Prozesse etc.) der zu modellierenden Welt vorgegeben. Ontologien werden herangezogen, um semantisch-konsistentes Verhalten von allen an einem Kommunikationsvorgang partizipierenden menschlichen und maschinellen Akteuren (d.h. Anwendern und Informationssystemen) zu gewährleisten und dienen dazu, den Austausch und das Teilen von Wissen zu erleichtern. Mittels Ontologien wird das gemeinsame Begriffsvokabular der interagierenden Akteure festgelegt, das für eine problembezogene Verständigung innerhalb einer bestimmten Anwendungsdomäne notwendig ist. Im Unterschied zu Informationsmodellen enthalten die in den Ontologien zum Einsatz kommenden Modellierungssprachen (wie z.B. OWL, RDF) Möglichkeiten zur Spezifikation einer formalen Semantik in Form von Integritäts- und Inferenzregeln.

Zwischen den folgenden Ontologie-Typen kann unterschieden werden:<sup>21</sup>

- **Struktur-Ontologien (Meta-Ontologien):** Mittels Struktur-Ontologien wird die für die Erstellung von Wissensmodellen verfügbare Sprache definiert. Sie spezifizieren die Ausdrucksmöglichkeiten einer Modellierungssprache. Von besonderer Bedeutung sind Struktur-Ontologien, wenn unterschiedliche Repräsentationsformalismen (Methoden und Sprachen) zur Beschreibung unterschiedlicher Teile eines Gesamtmodells verwendet werden. Die Integration dieser Teile kann dann über eine gemeinsame Struktur-Ontologie erfolgen.<sup>22</sup>
- **Basis-Ontologien:** Basis-Ontologien zeichnen sich vor allem durch eine hohe Wiederverwendbarkeit aus. Sie beschreiben fundamentale Konzepte und generische Aufgaben, die in fast jedem Anwendungsgebiet eine Rolle spielen und ausreichend standardisiert sind, um eindeutig beschrieben zu werden.
- **Domänen-Ontologien:** Domänen-Ontologien enthalten Wissen, das zur Beschreibung einer bestimmten Domäne notwendig ist.<sup>23</sup> In Domänen-Ontologien werden Informationseinheiten bezüglich Struktur, Strukturbestandteilen sowie Wissen über inhaltliche Zusammenhänge zwischen den Informationseinheiten beschrieben. Sie bilden die Grundlage für die Analyse eines Problembereichs. Darüber hinaus legt eine Domänen-Ontologie ein gemeinsames Vokabular zur Beschreibung der Domäne fest, wodurch die Akquisition von Domänenwissen erleichtert wird.

---

<sup>21</sup> Vgl. Wielinga, T. A., Schreiber, B. J. (1994), S. 197.

<sup>22</sup> Vgl. Gruber, T. R. (1993); Studer, R. et al. (1999), S. 5.; Fensel, D., Groenboom, R. (1997).

<sup>23</sup> Vgl. Eriksson, H. et al. (1994).

- **Methoden-Ontologien:** Im Unterschied zu Domänen-Ontologien beschreiben Methoden-Ontologien die relevanten Aspekte einer zu lösenden Aufgabe unabhängig von einer konkreten Domäne. Methoden-Ontologien können zur Beschreibung wiederverwendbarer Problemlösungsmethoden eingesetzt werden.<sup>24</sup>
- **Verhaltens-Ontologien:** Verhaltens-Ontologien werden zur Kommunikation eingesetzt, um die Konsistenz bestimmter Aktionsgrundmuster einzelner kooperativer menschlicher oder maschineller Akteure zu gewährleisten. Diese Ontologie-Variante betrifft somit das Reaktionsverhalten der Akteure und die zu realisierenden Aktionen. Die Verhaltens-Ontologie beschreibt folglich Wissen, das zur Lösung einer bestimmten Aufgabe in einer gegebenen Domäne verwendet wird.

Das hier vorgestellte Konzept der ontologie-basierten Attributierung von Informationsmodellen basiert im Wesentlichen auf Domänen-Ontologien. Wenngleich auch der Einsatz von Struktur-Ontologien zur Beschreibung von Modellierungssprachen denkbar wäre, wird aus wirtschaftlichen Gründen auf ihren Einsatz verzichtet. Stattdessen kommen in der Regel bereits vorliegende Metamodelle und Meta-Metamodelle zum Einsatz.

### 3 Überblick

Das Ziel der ontologie-basierten Attributierung von Informationsmodellen besteht darin, automatisierte semantische Analysen solcher Modelle zu ermöglichen. Auf diesem Weg soll es möglich sein, eine Vielzahl von semantischen Fragestellungen zu beantworten, z.B.:

- Wie ähnlich sind sich Informationsmodelle hinsichtlich ihrer semantischen Breite, d.h. wie groß ist die rekonstruierte Problemdomäne?
- Wie ähnlich sind sich Informationsmodelle hinsichtlich ihrer semantischen Tiefe, d.h. wie detailliert ist die Problemdomäne repräsentiert?
- Was sind effiziente und weniger effiziente Informationsmodell-Strukturen?
- Wie weit deckt ein Informationsmodell eine Problemdomäne ab?
- Welches Informationsmodell eignet sich besonders für ein konkretes Modellierungsziel?

Die hierzu erforderliche semantische Analyse von Informationsmodellen soll unabhängig von der konkreten Modellierungssprache sein, d.h. nicht an ein spezifisches Metamodelle gebunden sein. Aus diesem Grund bildet ein einheitliches Meta-Metamodelle die Grundlage für das Konzept der ontologie-basierten Attributierung von Informationsmodellen. Die Sprachkonstrukte der potenziell verwendeten Modellierungssprachen wer-

---

<sup>24</sup> Vgl. Gennari, J. et al. (1994).

den hierbei den im Meta-Metamodell definierten Sprachkonstrukten im Sinne einer Äquivalenz zugeordnet. Auf diese Weise wird expliziert, welche Elemente der Modellierungssprache mit welchen Elementen des Meta-Metamodells korrespondieren. Erfolgt diese Zuordnung für mehr als eine Modellierungssprache, kann eine einfache Übersetzung zwischen Modellen dieser Sprache vorgenommen werden.

Die semantische Interpretation der Informationsmodelle erfordert es, dass spezifiziert wird, welche Bedeutung den Elementen eines Informationsmodells zukommt. Hierzu werden diesen Elementen Konzepte einer Domänenontologie zugeordnet. Als Voraussetzung ist diese Domäne zu benennen und abzugrenzen und diese Ontologie zu entwickeln. Zur Integritätssicherung der nachfolgenden Zuordnung von Ontologiekonzepten zu Modellelementen werden im folgenden Schritt die einzelnen Ontologiekonzepte den Sprachkonstrukten des Meta-Metamodells zugeordnet, so dass deutlich wird, welche Ontologiebegriffe welchen syntaktischen Sprachkonstrukten entsprechen.

Auf Basis dieser Vorarbeiten kann eine Zuordnung von Konzepten der Ontologie zu den Elementen der betreffenden Informationsmodelle erfolgen. Als Ergebnis liegen Informationsmodelle vor, deren Bedeutung im Kontext der Konzepte der Ontologie interpretiert werden kann. Jedes Modell wird vor dem Hintergrund einer einheitlichen Sprache – nämlich der der Ontologie – verstanden und kann algorithmisch verarbeitet werden. Hierzu sind in erster Linie fortgeschrittene Ähnlichkeitsmaße für den Modellvergleich bzw. dem semantischen Vergleich der Ontologie und der Informationsmodelle erforderlich. Diese sind zweckbezogen zu entwickeln und anzuwenden.

Im Folgenden wird das Konzept im Detail erläutert.

## 4 Entwicklung eines syntaktischen Meta-Metamodells

In einem Metamodell werden die Ausdrucksmöglichkeiten, Sprachelemente und Konstruktionsregeln einer Modellierungssprache spezifiziert. In den unterschiedlichen Modellierungssprachen zur Informationsmodellierung finden sich syntaktische Elemente aufgrund des ähnlichen Aufgabenbereichs wieder, welche äquivalent zueinander sind. Mit dem „Entity Relationship Model“ (ERM)<sup>25</sup> und der „Unified Modeling Language“ (UML)<sup>26</sup> existieren z.B. zwei Sprachen, mit denen gleiche Realitätsausschnitte ohne Informationsverlust auf unterschiedliche Weise modelliert werden können. Darüber hinaus sind Modellierungssprachen aber auch auf eine bestimmte Sicht von Informationsmodellen spezialisiert, so dass es u.U. kaum Gemeinsamkeiten zwischen zwei Modellen gibt. Das ERM stellt so z.B. eine Struktur dar, wohingegen eine Ereignisgesteu-

---

<sup>25</sup> Vgl. Chen, P. P. (1976).

<sup>26</sup> Vgl. Spezifikation zu UML 2.0 (2003).

erte Prozesskette (EPK)<sup>27</sup> ein Verhalten verdeutlicht. Ein Meta-Metamodell ist in der Lage, diesen unterschiedlichen Gesichtspunkten bei dem Vergleich von Informationsmodellen Rechnung zu tragen.

Die äquivalenten syntaktischen Elemente verschiedener Modellierungssprachen stellen auf der höheren Abstraktionsebene ein Element eines Meta-Metamodells dar. Für die Zwecke dieses Konzeptes werden aus Kosten-/Nutzenüberlegungen nur die elementaren syntaktischen Elemente einer Sprache in das Meta-Metamodell aufgenommen. Auf diese Weise wird das Meta-Metamodell einfach gehalten und die Zuordnung von Elementen der Modellierungssprachen zu Elementen des Meta-Metamodells kann vergleichsweise einfach und schnell erfolgen. Der Preis für dieses Vorgehen ist ein Informationsverlust sowie ein damit einhergehender schwächerer Anspruch an das Meta-Metamodell.

Mit der Aufstellung eines Meta-Metamodells wird im Ergebnis erreicht, dass die ontologie-basierte Attributierung von Informationsmodellen nicht an eine bestimmte Modellierungssprache gebunden ist und somit sprachunabhängig ist. Außerdem bestehen neben dem semantischen Interesse an Informationsmodellen weiterhin die syntaktische Vergleichsmöglichkeiten verschiedener Modelle und sogar die Kombination beider Bereiche.

Die Zuordnung von Elementen eines Metamodells zu Elementen des Meta-Metamodell („Mapping“) kann wie folgt formalisiert werden:

*Modelle seien vereinfacht definiert als Menge  $M = \{S, R\}$ , wobei  $S$  eine Menge von Modellelementen und  $R$  eine Menge von Relationen zwischen diesen Elementen sei:  
 $R \subseteq S \times S$ .*

Gegeben seien:

$M$ : *ein Informationsmodell.*

$M^1$ : *ein Metamodell der Modellierungssprache, mit der  $M$  konstruiert ist.*

$M^2$ : *ein Meta-Metamodell, d.h. ein Metamodell der Sprache, mit der  $M^1$  konstruiert ist.*

Dabei gelten vereinbarungsgemäß folgende Mappings:

$m_1 : e \in M \rightarrow e^1 \in M^1$  *d.h. jedes Element  $e$  des Modells entspricht einem Element  $e^1$  des Metamodells.*

$m_2 : e^1 \in M^1 \rightarrow e^2 \in M^2$  *d.h. jedes Element  $e^1$  des Meta-Modells entspricht einem Element  $e^2$  des Meta-Metamodells.*

Geht man nun analog von einem zweiten Modell  $M'$  sowie einem zweiten Metamodell  $M'^1$  aus, das ebenfalls dem Meta-Metamodell  $M^2$  elementweise zugeordnet wird, dann

---

<sup>27</sup> Vgl. Scheer, A.-W. (2001), S. 125.

lassen sich Metamodellelemente  $e^1 \in M^1$  und  $e^1 \in M^1$  finden, die im Sinne des Meta-Metamodells syntaktisch äquivalent sind. Dies ist genau dann der Fall, wenn:

$$m_2(e^1) = m'_2(e^1)$$

### Beispiel

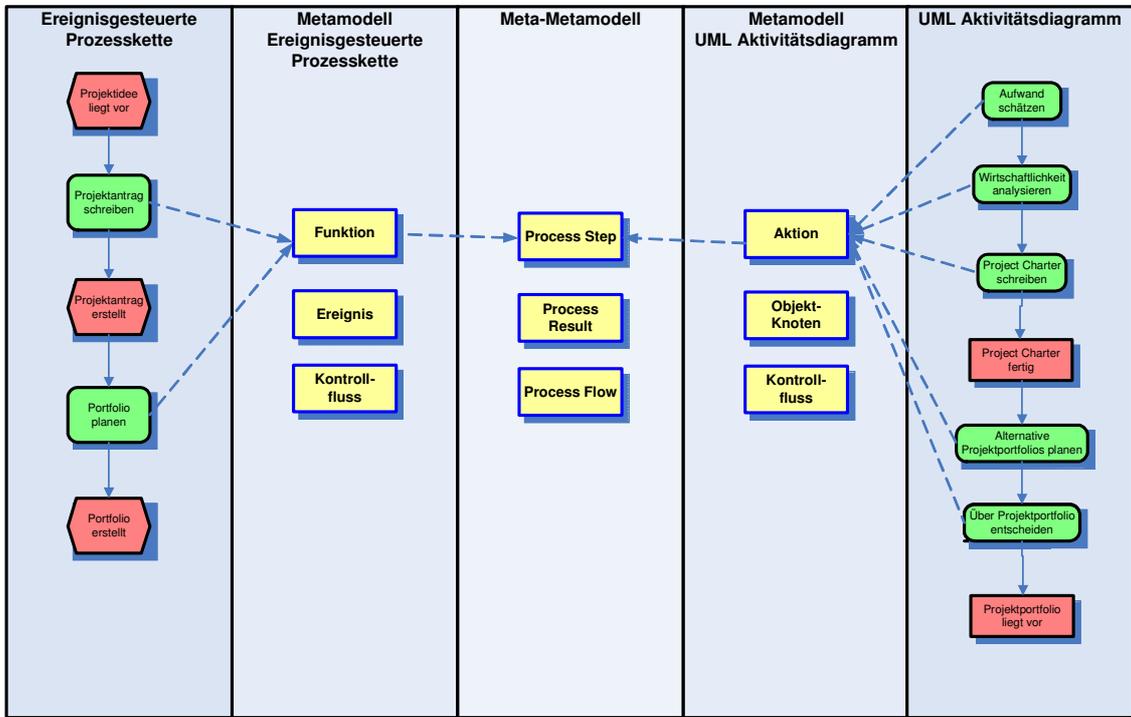


Abb. 3: Beispiel zu Meta-Metamodellen

In Abb. 3 liegen zwei Prozessmodelle desselben Realweltausschnittes vor, die unter Verwendung unterschiedlicher Modellierungssprachen konstruiert wurden. Das erste Modell ist eine ereignisgesteuerte Prozesskette, die nur drei Modellelementtypen verwendet: Funktionen, Ereignisse und den Kontrollfluss. Diese drei Typen sind elementare Bestandteile des entsprechenden Metamodells. Auf der anderen Seite ist derselbe Prozess als Aktivitätsdiagramm dargestellt. Hier sind die Elementtypen Aktion, Kontrollfluss und Objektknoten von zentraler Bedeutung und sind dementsprechend im Metamodell hinterlegt. Durch das Mapping zum Meta-Metamodell wird deutlich, dass die Modellelementtypen Aktion und Funktion syntaktisch äquivalent sind.

Aus Gründen der Übersichtlichkeit wird in diesem Beispiel nur das Mapping von Funktionen/Aktionen/Process Steps dargestellt. Analog können andere Elementtypen ebenfalls zugeordnet werden.

## 5 Entwicklung einer problemangemessenen Domänenontologie

Die semantische Analyse von Informationsmodellen erfordert die Zuhilfenahme von Ontologien, die es vorab zu entwickeln gilt. Die Problemangemessenheit einer Ontologie ist gegeben, wenn alle fachlichen Aspekte der potenziell zu attributierenden Modelle von der Ontologie abgedeckt werden. Um dieses zu gewährleisten, bietet sich der Rückgriff auf Standards, Normen und Referenzmodelle an, die im Zusammenhang mit Informationsmodellen für die Problemdomäne entwickelt worden sind. Der Detaillierungsgrad dieser Aufgabe wird dabei von wirtschaftlichen Gesichtspunkten bestimmt.<sup>28</sup> Eine zu weit fortgeschrittene Detaillierung ist zu aufwendig und bildet eventuell Aspekte ab, die keinen Nutzen entfalten, da sie nur in wenigen Modellen abgebildet sind. Eine zu grobe Detaillierung lässt dagegen u.U. nur wenig aussagekräftige Analysen zu. Die Problemangemessenheit einer Ontologie stellt insofern schon für sich genommen eine offene Forschungsfrage dar.

Die zu entwickelnde Ontologie hat allgemein aus einem Vokabular mit den Konzepten der Problemdomäne und aus einer Menge von logischen Aussagen über diese Begriffe und ihre Zusammenhänge zu bestehen.<sup>29</sup> Dieses übliche Vorgehen wird für die semantische Attributierung von Informationsmodellen um einen weiteren Schritt ergänzt. Jedes Konzept der Ontologie wird einem syntaktischen Element des Meta-Metamodells zugeordnet. So wird deutlich, ob es sich bei einem Konzept um eine Funktion, eine Datenstruktur oder um etwas anderes handelt.

$$m_3 : k \in K \rightarrow e^2 \in M^2$$

Die Funktion  $m_3$  bildet ein Konzept  $k$  aus der Konzeptmenge  $K$  der Ontologie auf ein Element  $e^2$  des Meta-Metamodells  $M^2$  ab.

Zusätzlich können mit Hilfe der Ontologie weitergehende Integritätsregeln definiert werden, die helfen, die Wahrscheinlichkeit von späteren fehlerhaften Attributierungen von Informationsmodellen zu vermeiden.

---

<sup>28</sup> Schütte zählt die Wirtschaftlichkeit zu den Grundsätzen ordnungsmäßiger Modellierung. Vgl. Schütte, R. (1998), S. 127 ff.

<sup>29</sup> Vgl. Mädche, A. et al. (2001), S. 393.

## Beispiel (fortgeführt)

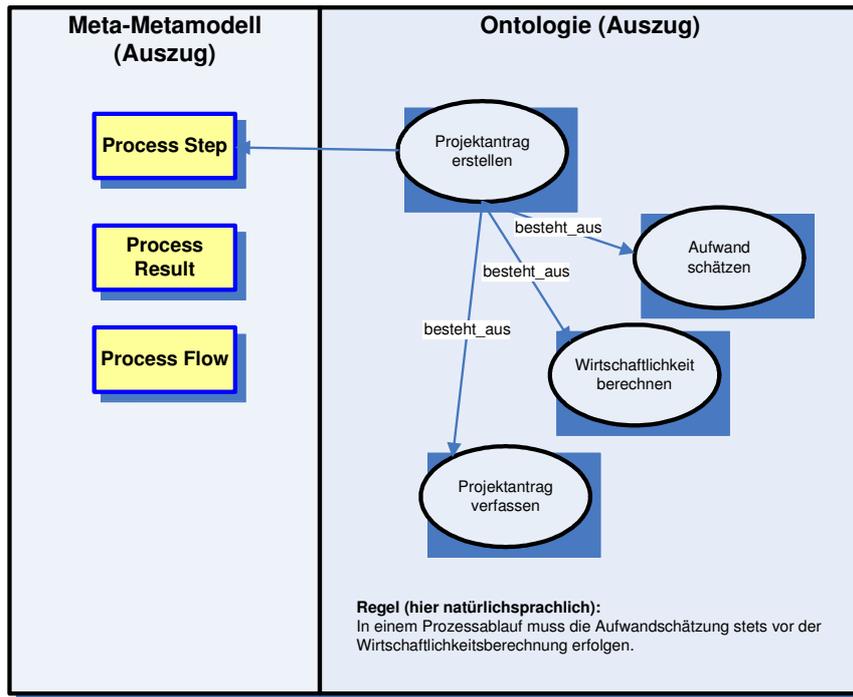


Abb. 4: Beispiel zur Ontologie-Konstruktion

Im fortgesetzten Beispiel (Abb. 4) umfasst die Ontologie (auszugsweise) vier Konzepte, die hierarchisch angeordnet sind. Das Konzept „Projektantrag erstellen“ ist hier den übrigen drei Konzepten übergeordnet. Diese Anordnung verdeutlicht eine hierarchische Zerlegung der Gesamtaufgabe. Darüber hinaus ist das Konzept „Projektantrag erstellen“ dem Meta-Metamodell-Element „Process Step“ zugeordnet. Auf diese Weise kann es später nur Modellelementen zugeordnet werden, die einen Prozessschritt darstellen. Als weitere Integritätsregel wurde definiert (hier aus Gründen des Umfangs natürlichsprachlich), dass in einem Prozessmodell eine Aufwandschätzung stets vor einer Wirtschaftlichkeitsberechnung zu erfolgen hat.

Das skizzierte Beispiel macht bereits deutlich, dass sich die Funktionen des Meta-Metamodells sowie der Ontologie überlappen. Aus diesem Grund wäre es denkbar, die Ontologie im Sinne einer Strukturontologie auch um syntaktische Konzepte und entsprechende Regeln anzureichern und somit das Meta-Metamodell überflüssig zu machen.

## 6 Semantische Attributierung von Informationsmodellen

Durch das Aufstellen einer Ontologie und die Zuordnung von Ontologiekonzepten zu den Elementen eines Meta-Metamodells sind die Voraussetzungen gegeben, um ein Informationsmodell mit semantischen Attributen zu versehen. Es ist nun möglich, das Element  $e$  eines spezifischen Informationsmodells  $M$  einem einzelnen Konzept  $k \in K$  einer Ontologie  $O$  zuzuordnen. Dies beschreibt die Funktion  $m_4$  formal:

$$m_4 : e \in M \rightarrow k \in K$$

wobei die Bedingung

$$m_3(k) = m_2(m_1(e))$$

gelten muss, d.h. es sind nur solche Ontologiekonzepte zuzuordnen, die syntaktisch aufgrund des Meta-Metamodells erlaubt sind. Je vollständiger die Modellelemente entsprechenden Ontologieelementen zugeordnet werden können, desto höher ist die Güte der Attributierung.

Zusätzlich zur Zuordnung von Ontologiekonzepten kann der Ansatz der ontologiebasierten Attributierung dahingehend erweitert werden, dass einzelne Modellelemente  $e \in M$  mit einem zusätzlichen Merkmalsvektor versehen werden, der Informationen enthält, die für die Analyse des Informationsmodells notwendig oder hilfreich sind. So wäre bspw. denkbar, dass zu einem Prozessschritt Kosteninformationen oder Durchlaufzeiten erfasst und zugeordnet werden. Dies kann wie folgt formalisiert werden:

Einzelnen Modellelementen  $e$  kann wie folgt ein Merkmalsvektor  $b$  zugeordnet werden:

$$m_5 : e \in M \rightarrow b = (b_1, \dots, b_n)$$

Hierzu lassen sich Integritätsbedingungen definieren, die für ein Ontologiekonzept oder aber ein Meta-Metamodell-Element festlegen, welche Merkmalsvektoren zulässig sind. Ein Informationsmodell ist nach der Zuordnung von Ontologiekonzepten und Merkmalsvektoren nicht länger ein Gebilde, welches allein formalen Konstruktionsregeln gehorcht. Innerhalb der Modellstruktur ist mit Hilfe der Ontologie die Bedeutung und Logik einzelner Modellelemente abgebildet. So lassen sich bspw. die bei einem automatischen Vergleich von Informationsmodellen typischen Abstraktions-, Namens-, Detaillierungs- und Typkonflikte lösen (vgl. Abb. 5). Das Ergebnis stellt die Grundlage für eine semantische Analyse von Informationsmodelle dar und eröffnet neue Einsatzfelder für diese Modelle.

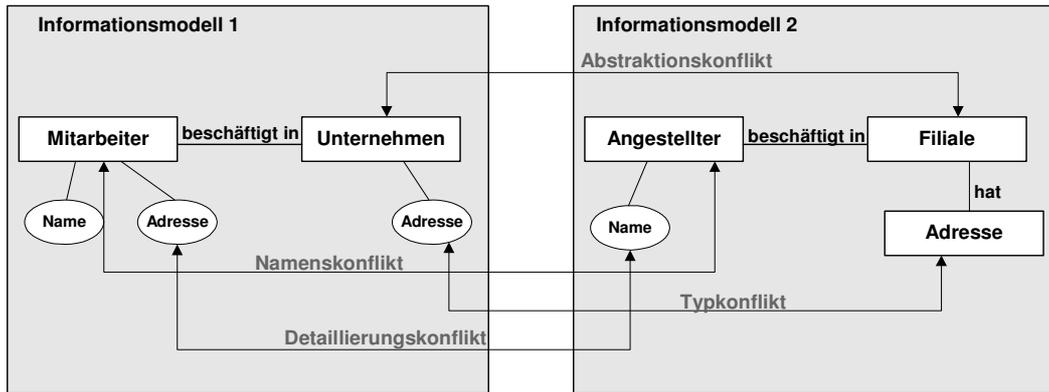


Abb. 5: Konflikttypen beim Vergleich von Informationsmodellen

Beispiel (fortgeführt)

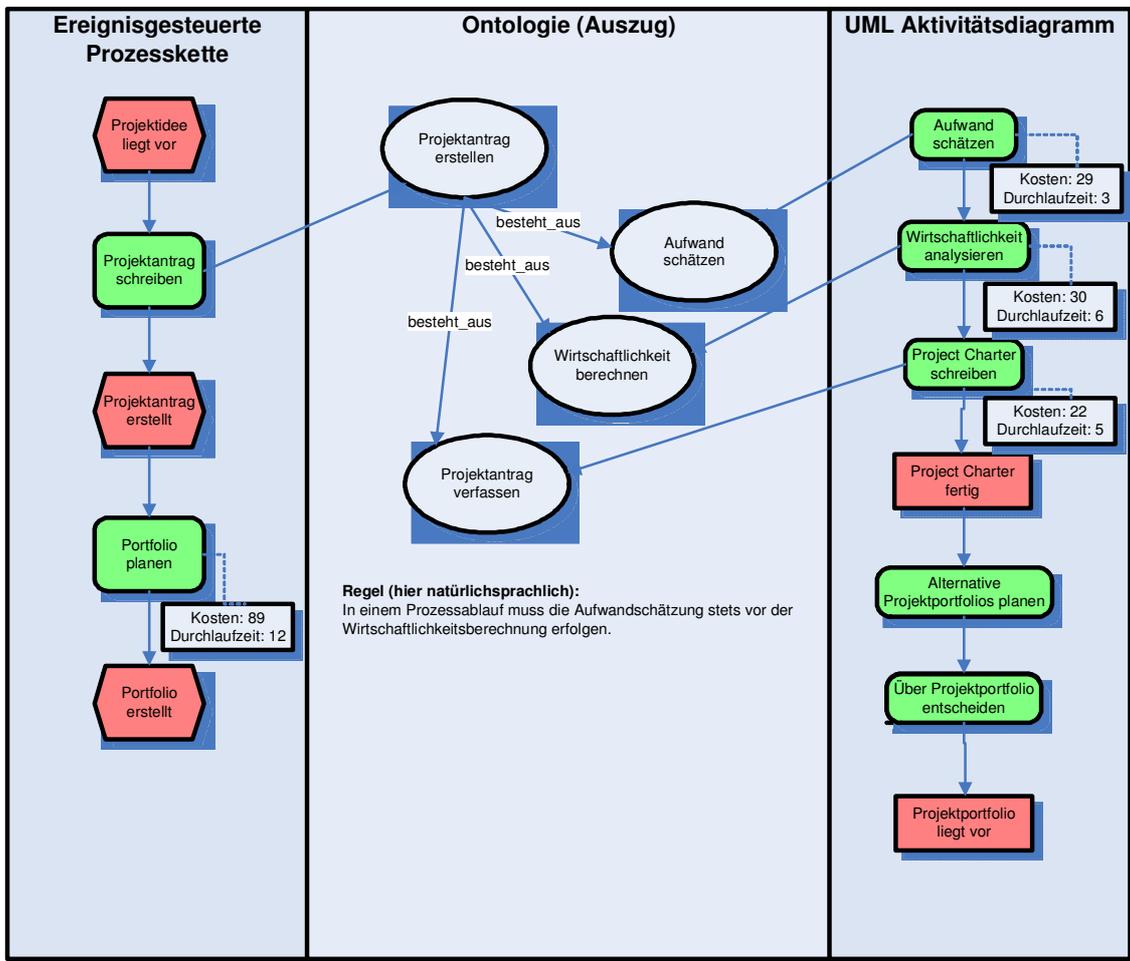


Abb. 6: Semantische Attributierung auf Basis von Ontologien

Im fortgesetzten Beispiel (vgl. Abbildung 6) ist aufgrund der Zuordnung von Konzepten der Ontologie zu Modellelementen nun ersichtlich, dass der Prozessschritt „Projektantrag schreiben“ des ersten Modells identisch mit drei unterschiedlichen Prozessschritten im zweiten Modell ist. Hieraus lässt sich bspw. auch ableiten, dass die Informationen zu Kosten und Durchlaufzeiten des zweiten Modells entsprechend zu aggregieren sind, um sie mit dem ersten Modell vergleichen zu können.

## 7 Automatisierter Vergleich von Informationsmodellen

Um zwei Informationsmodelle miteinander vergleichen zu können, sind zweckbezogene Metriken zu definieren. Dies können grundlegende Metriken sein, die z.B. die Gesamtzahl von Elementtypen in einem Modell berechnen oder die durchschnittliche Anzahl von Elementen pro Elementtyp zählen. Zur Charakterisierung von Modellen gibt es darüber hinaus Kohäsionsmetriken, die das Beziehungsgeflecht zwischen den Elementen eines Modells messen. Weiterhin existieren Komplexitätsmetriken als Grad für die Komplexität eines Modells. Metriken können auch auf den in Abschnitt 6 eingeführten Merkmalsvektoren operieren.

Nachdem die Metriken definiert wurden, sind sie zu Merkmalsvektoren für das Gesamtmodell zu kombinieren. Die Ähnlichkeit zwischen den Merkmalsvektoren zweier Informationsmodelle kann wiederum durch Distanzmaße, wie z.B. die Euklidische Distanz oder die City-Block-Distanz, ausgedrückt werden. In diesem Zusammenhang ist zu überprüfen, ob unterschiedliche Größenordnungen in den Merkmalsvektor einfließen. Es ist zu entscheiden, ob durch Normierung und Skalierung von Metriken Verzerrungen im Merkmalsraum auszugleichen sind.

### 7.1 Metriken

Einfache automatisierte Analysen der Informationsmodelle auf Basis der semantischen Attributierung sind sofort möglich. Um jedoch fortgeschrittene automatisierte Analysen wie z.B. zur semantischen Ähnlichkeit von Informationsmodellen zu ermöglichen, ist es notwendig und sinnvoll, zweckbezogen zusätzliche Metriken zu definieren. Dann können bspw. Distanz- und Ähnlichkeitsmaße zum Einsatz kommen, die auf diesen Metriken operieren.

Metriken lassen sich wie folgt klassifizieren:<sup>30</sup>

- **Grundlegende Metriken:** Diese Metriken berechnen die Gesamtzahl bestimmter Elementtypen (Klassen, Attribute, etc.) innerhalb eines Modells bzw. die

---

<sup>30</sup> Vgl. Yao et al. (2005), S. 108 ff.

durchschnittliche Anzahl pro Elementtyp (z.B. die durchschnittliche Anzahl der Attribute pro *Klasse* innerhalb eines Modells).

- **Gewichtete, additive Metriken:** Diese Metriken verknüpfen Metriken zu neuen Metriken, indem sie einzelne Metriken gewichten. Häufig wird eine additive Verknüpfung gewählt.
- **Kohäsionsmetriken:** Kohäsionsmetriken messen das „Beziehungsgeflecht“ zwischen den einzelnen Elementen eines Informationsmodells.<sup>31</sup>
- **Komplexitätsmetriken:** Komplexitätsmetriken dienen als Maß für den Komplexitätsgrad eines Informationsmodells; gemessen werden hierbei bspw. in objektorientierten Modellen Generalisierungen, Polymorphie, die Anzahl der Methodenaufrufe oder Kapselungen.

In den letzten Jahren wurde eine Vielzahl von Metriken zur Bewertung und zum Vergleich von (objektorientierten) Informationsmodellen entwickelt. Chidamber und Kemerer<sup>32</sup> haben einen Satz von Metriken definiert, auf denen eine Vielzahl weiterer Metriken von anderen Autoren basiert.<sup>33</sup> Tabelle 1 fasst eine Auswahl der wichtigsten grundlegenden Metriken zusammen.<sup>34</sup>

---

<sup>31</sup> Vgl. Yao et al. (2005), S. 108.

<sup>32</sup> Vgl. Chidamber, S.R., Kemerer, C.F. (1994) und Chidamber, S.R. et al. (1998).

<sup>33</sup> Vgl. Lorenz, M., Kid, J. (1994); Reeves, B.N., Hayashi, Y. (1999); Rosenberg, L.H. (1998).

<sup>34</sup> Vgl. Yao et al. (2005), S. 108 ff.

## Grundlegende Metriken

- Total Number of Classes (**TNC**): Anzahl der Klassen
- Total Number of Children (**TNOC**): Anzahl der Unterklassen
- Total Number of Class Attributes (**TNCA**): Anzahl der verwendeten Attribute
- Average Depth of Inheritance Trees of a Class (**DIT**): Durchschnittliche Tiefe der Vererbungshierarchie einer Klasse
- Average Number of Children of a Class (**NOC**): Durchschnittliche Anzahl der Unterklassen einer Klasse
- Average Number of Attributes per Class (**NCA**): Durchschnittliche Anzahl der Attribute einer Klasse
- Maximal Depth of Inheritance Tree (**MDIT**): Maximale Tiefe der Vererbungshierarchie von der Wurzelklasse zu einer Endklasse
- Total Number of Relationships (**TNOR**): Anzahl der Relationen auf einer Hierarchieebene
- Total Number of Roles (**TOR**): Anzahl der Rollen
- Total Number of Inheritance Relationships (**TNIR**): Anzahl der Relationen zwischen über- und untergeordneten Hierarchieebenen
- Total Number of Actors in a Model (**NACM**): Anzahl der Akteure in einem Informationsmodell
- Average Number of Relationships per Class (**NRC**): Durchschnittliche Anzahl der Relationen pro Klasse
- Average Number of Inheritance Relationships per Class (**NIR**): Durchschnittliche Anzahl der Relationen zwischen über- und untergeordneten Hierarchieebenen
- Total Number of Objects in a Model (**NOM**): Anzahl der Objekte in einem Modell
- Total Number of Use Cases in a Model (**NUM**): Anzahl der Anwendungsfalldiagramme in einem Modell

**Tab. 1: Ausgewählte grundlegende Metriken zum Vergleich von Ontologien**

Tabelle 2 fasst ausgewählte Kohäsions- und Komplexitätsmetriken zusammen.<sup>35</sup> Auch diese Metriken zielen im Wesentlichen auf objektorientierte Informationsmodelle ab.

---

<sup>35</sup> Vgl. Van Elsuwe, H., Schmedding, D. (2003), S. 24 ff.

Typ	Metrik
<b>Kohäsionsmetriken</b>	<ul style="list-style-type: none"> <li>- Number of Root Classes (<b>NoR</b>): Anzahl der Wurzelklassen</li> <li>- Number of Leaf Classes (<b>NoL</b>): Anzahl der Klassen auf der untersten Hierarchieebene</li> <li>- Average Depth of Inheritance Tree of Leaf Nodes (<b>ADIT-LN</b>): Durchschnittliche Tiefe der Vererbungshierarchie von allen Klassen auf den untersten Hierarchieebenen</li> <li>- Average Number of Actors associated with a Use Case (<b>ANACU</b>): Durchschnittliche Anzahl der Akteure, die mit einem Anwendungsfall assoziiert werden</li> <li>- Total Number of Aggregations in a Model (<b>NAGM</b>): Anzahl der Aggregationen in einem Informationsmodell</li> <li>- Average Number of Superclasses of a Class (<b>NSUBC</b>): Durchschnittliche Anzahl von Oberklassen zu einer Klasse in einem Informationsmodell</li> <li>- Total Number of Associations (<b>TNA</b>): Anzahl der Assoziationen</li> <li>- Lack of Cohesion of Methods (<b>LCOM</b>): Der Wert von LCOM berechnet sich aus der Anzahl an Paaren <math> P </math> von Methoden in einer Klasse, die keine gemeinsame Instanzvariable haben, minus der Anzahl an Paaren <math> Q </math> von Methoden in einer Klasse, die eine gemeinsame Instanzvariable haben. Ein größerer Wert von LCOM steht für eine geringere Kohäsion (wenn <math> P  &lt;  Q </math>, wird der Wert von LCOM auf 0 gesetzt).</li> <li>- Reuse Ratio (<b>RR</b>): Die Reuse Ratio ist der Quotient aus <math>NoR</math> dividiert durch <math>TNC</math></li> <li>- Specialization Ratio (<b>SR</b>): Die Specialization Ratio ist der Quotient aus <math>TNOC</math> und <math>NoR</math>.</li> </ul>
<b>Komplexitätsmetriken</b>	<ul style="list-style-type: none"> <li>- Informative Größe (<b>INF</b>): Zur Bestimmung der informativen Größe eines Informationsmodells wird den Hierarchieebenen eine Gewichtung zugeordnet und die Anzahl der Elemente einer Hierarchieebene gezählt. Mit Hilfe einer einfachen Formel, in der die Anzahl der Modellelemente (<math>E_i</math>) einer Ebene mit der Gewichtung (<math>G_i</math>) dieser Ebene multipliziert wird, lässt sich die informelle Größe INF einer Ontologie berechnen, wobei <math>n</math> die Anzahl der Ebenen der Elementhierarchie ist: <math display="block">INF = \sum_{i=1}^n E_i G_i</math> </li> <li>- Method Hiding Factor (<b>MHF</b>): Eine Metrik zum Messen des Ausmaßes von Information Hiding in Methoden</li> <li>- Attribute Hiding Factor (<b>AHF</b>): Eine Metrik zum Messen des Ausmaßes von Information Hiding durch lokal definierte Attribute</li> <li>- Method Inheritance Factor (<b>MIF</b>): Eine Metrik zum Messen des Ausmaßes von Information Hiding durch lokal definierte Methoden</li> </ul>

**Tab. 2: Kohäsions- und Komplexitätsmetriken zum Vergleich von Ontologien**

Die hier skizzierten Metriken sind im Hinblick auf das in dieser Arbeit vorgestellte Konzept der semantischen Attributierung von Informationsmodellen weiterzuentwickeln.<sup>36</sup> Bspw. sind die folgenden Ansätze denkbar:

- Das in der Domänenontologie codierte und den Informationsmodellelementen zugeordnete Konzeptvokabular kann direkt in einem Merkmalsvektor codiert

<sup>36</sup> Vgl. Abschnitt 10.

werden. Eine sehr einfache Codierung wäre z.B. die binäre Codierung mit 0 (Konzept  $k$  ist nicht vorhanden) und 1 (Konzept  $k$  ist vorhanden).

- Die Tiefe, d.h. Detaillierung eines Informationsmodells, kann durch die Anzahl der Referenzierungen auf in der Konzepthierarchie tief stehende Konzepte zum Ausdruck gebracht werden.
- Die Breite, d.h. die Umfänglichkeit eines Informationsmodells, kann durch die Anzahl der abgedeckten Teiläste der Konzepthierarchie gemessen werden.

### Beispiel (fortgeführt)

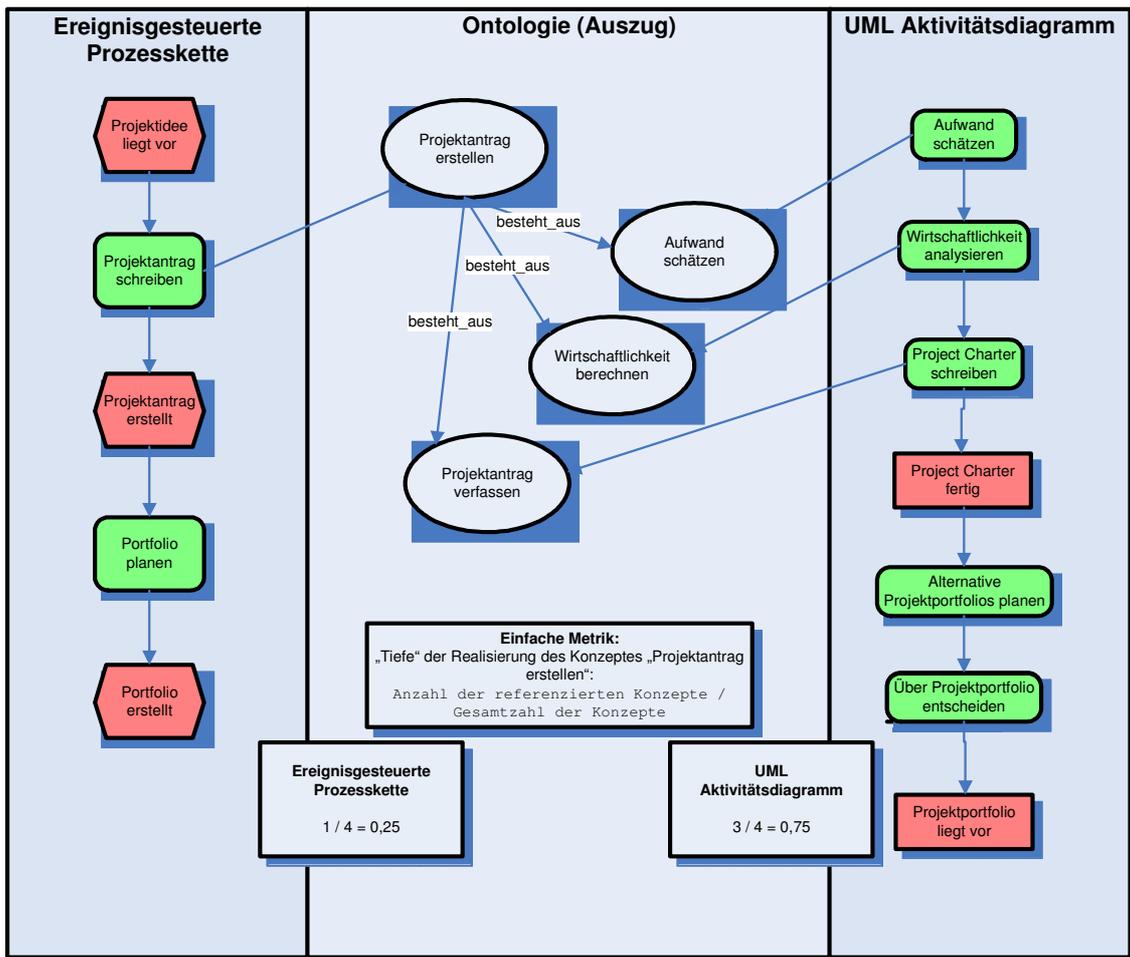


Abb. 7: Metrik zum Informationsmodellvergleich

Im fortgesetzten Beispiel (Abbildung 7) kommt eine sehr einfache Metrik zum Modellvergleich zum Einsatz. Mit dieser Metrik kann gemessen werden, wie detailliert ein Konzept in einem Informationsmodell repräsentiert ist. Die Metrik besteht aus dem Quotienten aus referenzierten (Sub-)Konzepten und der Anzahl vorhandener (Sub-)Konzepte. In Hinblick auf das Konzept „Projektantrag erstellen“ ergibt sich hier für die ereignisgesteuerte Prozesskette ein Wert von 0,25, weil nur eines von vier Konzepten

referenziert wird. In Hinblick auf das Aktivitätsdiagramm ergibt sich ein Wert von 0,75, weil drei von vier Konzepten referenziert werden.

Wie leicht zu erkennen ist, kann diese Analyse vollständig automatisiert erfolgen. Voraussetzung ist lediglich, dass die beiden Informationsmodelle entsprechend semantisch attribuiert sind. Dann spielen auch Namens-, Typ- und Detaillierungsprobleme keine Rolle mehr.

## 7.2 Distanz- und Ähnlichkeitsmaße

Die Ergebnisse der Anwendung von Metriken sowie andere den Modellelementen zugewiesene Merkmale (siehe Abschnitt 6) können zu einem Merkmalsvektor kombiniert werden, auf dem dann Distanzmaße wie bspw. die *Euklidische Distanz* oder die *City-Block-Distanz* operieren können. Auf diese Weise lassen sich Informationsmodelle mehrdimensional miteinander vergleichen. Denkbar sind auch spezialisierte, anwendungsfallbezogene Distanz- und Ähnlichkeitsmaße, der Einsatz von Künstlichen Neuronalen Netzen oder der Einsatz von hybriden Verfahren, d.h. eine sinnvolle Kombination von statistischen Verfahren (z.B. Diskriminanz- und Clusterverfahren), Neuronalen Netzen und/oder genetischen Algorithmen. Künstliche Neuronale Netze können bspw. die Merkmalsvektoren als Input verwenden und dann selbstorganisierend eine Klassifikation von Informationsmodellen vornehmen oder anhand der Merkmalsvektoren Ähnlichkeitsmaße zwischen jeweils zwei Informationsmodellen ermitteln. In einem N-dimensionalen Vektorraum ( $N = \text{Merkmalsanzahl}$ ) mit den Merkmalsvektoren  $\vec{x}$  und  $\vec{y}$ , welche die zwei Informationsmodelle *A* und *B* durch eine festgelegte Anzahl von Merkmalen bzw. Metriken beschreiben, liegen ähnliche Merkmalsvektoren im Vektorraum nahe beieinander. Die Distanz zwischen den Merkmalsvektoren (und somit die Ähnlichkeit der Informationsmodelle) lassen sich auf der Basis einer Vielzahl an Distanzmaßen messen. Tabelle 3 fasst ausgewählte Distanzmaße zusammen, die hierzu herangezogen werden können.

Distanzmaß	Formel
L <sub>1</sub> -Norm (Manhattan Distanz, City-Block-Distanz)	$L_1(\vec{x}, \vec{y}) = \sum_{i=1}^N  x_i - y_i $
L <sub>2</sub> -Norm (Euklidische Distanz)	$L_2(\vec{x}, \vec{y}) =  \vec{x} - \vec{y}  = \sqrt{\sum_{i=1}^N (x_i - y_i)^2}$
Cosinus-Distanz	$\cos(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{ \vec{x}  \cdot  \vec{y} } = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2} \cdot \sqrt{\sum_i y_i^2}}$
Die L <sub>1</sub> -Norm und die L <sub>2</sub> -Norm sind Spezialfälle des <i>Minkowsky</i> -Maßes	$D_{minkowsky}(\vec{x}, \vec{y}) = L_r(\vec{x}, \vec{y}) = \left( \sum_{i=1}^N (x_i - y_i)^r \right)^{\frac{1}{r}}$
Camberra-Distanz	$D_{camberra}(\vec{x}, \vec{y}) = \sum_{i=1}^N \frac{ x_i - y_i }{ x_i + y_i }$
Chebychev-Distanz	$D_{chebychev}(\vec{x}, \vec{y}) = \max_{i=1}^N  x_i - y_i $

Tab. 3: Distanzmaße zum Vergleich von Informationsmodellen mittels Merkmalsvektoren

### 7.3 Normierung und Skalierung

Ziel einer Normierung ist, unterschiedliche Merkmale der Informationsmodelle auf den gleichen Wertebereich zu transformieren und auf diese Weise Verzerrungen im Merkmalsraum auszugleichen.

Der Grund dafür liegt in den oftmals großen Spannweiten einzelner Merkmale, die aufgrund verwendeter Distanzmetriken oder Gewichtungen stärker in das Ergebnis eines Vergleichs einfließen würden als Merkmale mit kleinen Spannweiten (wie z.B. bei den binär codierten Merkmalen).

#### Normierung

Mit der folgenden Gleichung lassen sich bspw. Merkmalsausprägungen  $x_{nj}$  in normierte Werte  $x_{nj}^*$  aus dem Intervall  $[0,1]$  transformieren:<sup>37</sup>

$$x_{nj}^* = \frac{x_{nj} - x_{n,\min}}{x_{n,\max} - x_{n,\min}} \quad \forall n = 1, \dots, N; \forall j = 1, \dots, J$$

Dabei bezeichne  $N$  die Anzahl der Merkmale,  $J$  die Anzahl der betrachteten Informationsmodelle und  $x_{n,\min}$  die kleinste Ausprägung von Merkmalen  $n$  sowie  $x_{n,\max}$  die größte Ausprägung von Merkmalen  $n$  ( $n = 1, \dots, N$ ) für ein Informationsmodell  $j$  ( $j = 1, \dots, J$ ).

#### Skalierung

Lineare Skalierung ist die lineare Transformation der Eingabedaten auf eine andere Skala. Hierzu wird von einem normierten Wert  $x_{nj}^*$  ein Wert *offset* subtrahiert und das Ergebnis durch einen Wert *quotient* dividiert.

<sup>37</sup> Vgl. Zimmermann, H.-J. et al. (1995), S. 11.

Eine häufig verwendete Skalierung ist die lineare Transformation auf einen bestimmten Wertebereich  $|x_u, x_o|$  mit  $x_u < x_o$  durch:<sup>38</sup>

$$x_{nj}^{skal} = \frac{x_{nj}^* - \text{offset}}{\text{quotient}},$$

$$\text{mit } \text{quotient} = \frac{x_{n,max} - x_{n,min}}{x_o - x_u} \text{ und } \text{offset} = (x_{n,max} - \text{quotient} * x_o) = (x_{n,min} - \text{quotient} * x_u).$$

Mit der oben beschriebenen Normierung und der linearen Skalierung der Werte von Merkmalsvektoren lassen sich die unterschiedlichen Größenordnungen einzelner Variablen wie z.B. von binär codierten Merkmalen und Merkmalen mit kontinuierlichen Werten wie die Anzahl der Hierarchieebenen einer Ontologie angleichen. Bei der Verwendung von Distanzmetriken oder dem Einsatz von Künstlichen Neuronalen Netzen sollten somit Normierung und Skalierung der Merkmalsvektoren vorangeschaltet werden, um Verzerrungen in den Ergebnissen eines Vergleichs von Ontologien zu vermeiden.

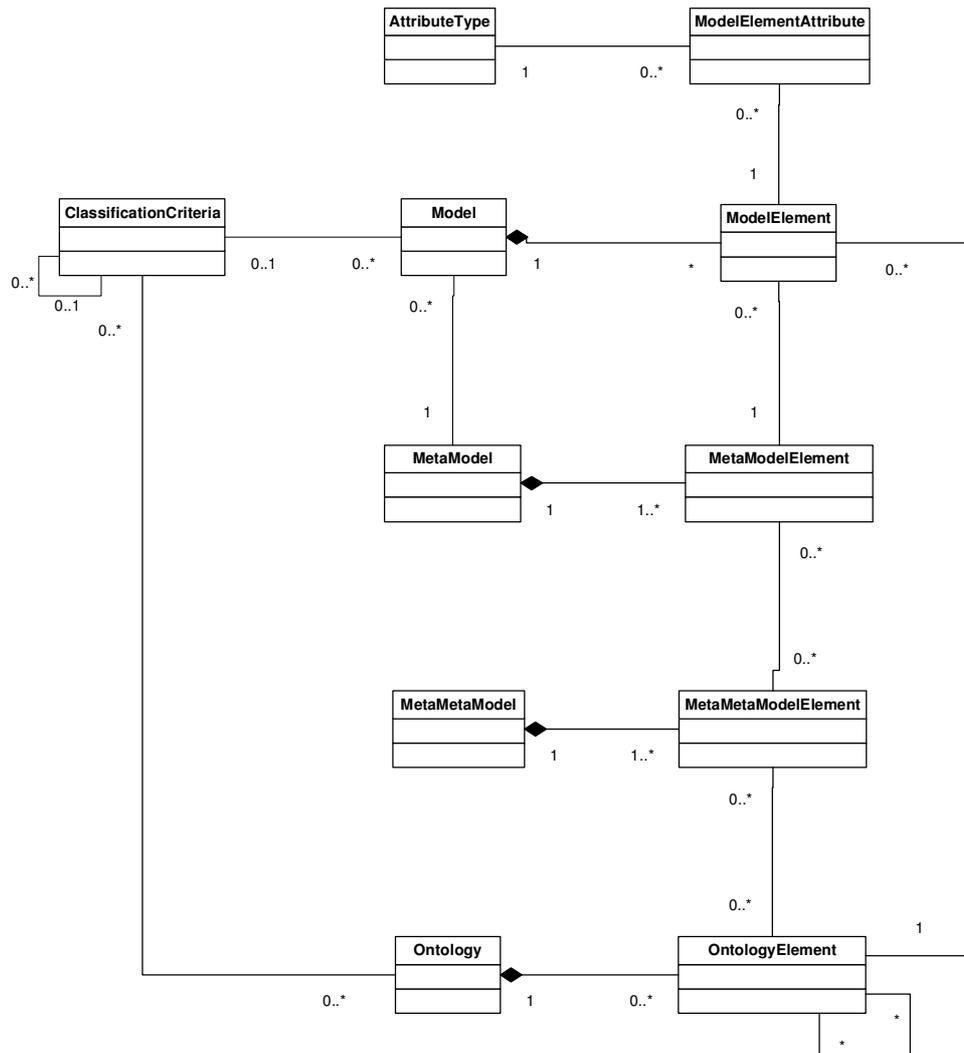
## 8 Architektur eines allgemeinen Informationsmodell-Datenbanksystems

Für eine ontologie-basierte Attributierung von Informationsmodellen werden die entwickelte Ontologie, die Metamodelle der verwendeten Modellierungssprachen und das Meta-Metamodell zusammen mit den Informationsmodellen in einer Modelldatenbank abgebildet (schematisch dargestellt in Abbildung 8 als UML-Klassendiagramm).

Ein einheitliches Meta-Metamodell wird aufgestellt, damit die semantische Attributierung unabhängig von einer konkreten Modellierungssprache ist. Die verschiedenen Sprachkonstrukte potenzieller Modellierungssprachen werden den Sprachkonstrukten des Meta-Metamodells im Sinne einer Äquivalenz zugeordnet. Dadurch wird deutlich, welche Konstrukte unterschiedlicher Modellierungssprachen äquivalent zueinander sind. Die Ontologiekonzepte werden ebenfalls den Sprachkonstrukten des Meta-Metamodells zugeordnet, um die Integrität der Zuordnung von Ontologiebegriffen zu Modellelementen sicherzustellen. Dadurch wird ersichtlich, welche Ontologiekonzepte welchen syntaktischen Konstrukten entsprechen.

Im Anschluss an diese Vorarbeiten können die Elemente eines Prozessmodells den entsprechenden Konzepten der Ontologie zugeordnet werden, die syntaktisch aufgrund des Meta-Metamodells erlaubt sind. Elemente unterschiedlicher Modelle, die einen gleichen Sachverhalt mittels verschiedener Ausdrücke beschreiben, werden so demselben Konzept der Ontologie zugeordnet.

Nachdem für die Modellelemente Entsprechungen in der Ontologie gefunden sind, liegen Informationsmodelle vor, deren Bedeutung im Kontext der Ontologie interpretiert werden kann. Durch die semantische Attributierung ist ein Informationsmodell somit nicht länger nur ein Gebilde, welches allein formalen Konstruktionsregeln gehorcht und von Individuen interpretiert werden muss. Innerhalb der Modellstruktur sind mit Hilfe der Ontologie die Bedeutung und die Logik der einzelnen Modellelemente abgebildet.

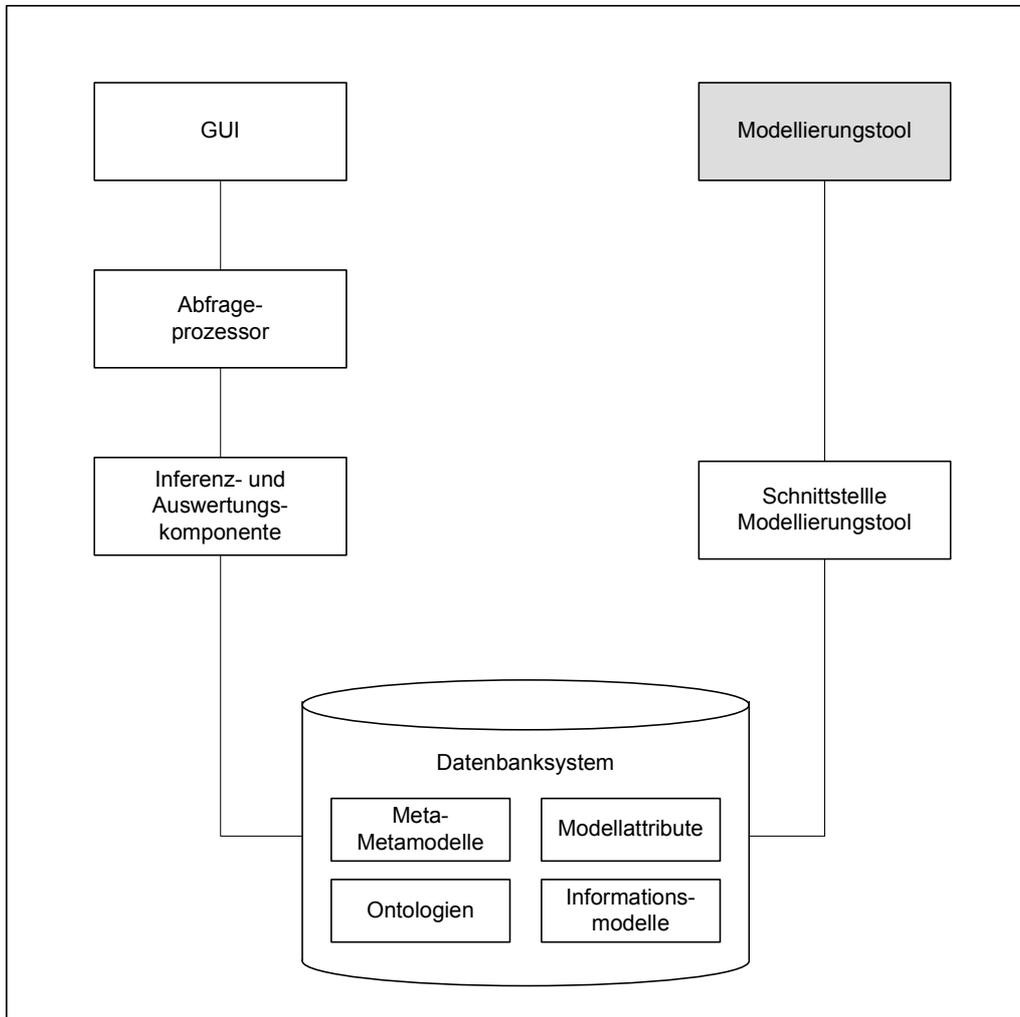


**Abb. 8: Ontologie-basierte Attributierung als UML-Diagramm**

Das in dieser Arbeit beschriebene Konzept einer ontologie-basierten Attributierung von Informationsmodellen ist nur durch eine adäquate Softwareunterstützung effizient einsetzbar. Die Attributierung vorhandener Informationsmodelle und die sich anschließende Auswertung sind arbeitsintensive Prozesse, die durch entsprechende Werkzeuge ver-

<sup>38</sup> Vgl. Zimmermann, H.-J. et al. (1995), S. 11.

einfacht werden können. Nachstehend werden deshalb die Grundzüge eines allgemeinen Informationsmodell-Datenbanksystems skizziert, in dem das Konzept umgesetzt ist (vgl. Abbildung 9).



**Abb. 9: Grundzüge der Architektur eines Informationsmodell-Datenbanksystems**

Die technische Grundlage für das Informationsmodell-Datenbanksystem bildet die Modelldatenbank, bei der es sich um ein klassisches relationales Datenbankmanagementsystem handeln kann. In dieser Modelldatenbank werden Meta-Metamodelle, Ontologien und attributierte Informationsmodelle abgelegt.

Da ein alleiniger Zugriff auf die Datenbank mit einer typischen Datenbankabfragesprache wie SQL nicht ausreicht, wird der Modelldatenbank eine Inferenz- und Auswertungskomponente vorgeschaltet, die komplexe Analysen auf Basis der zuvor beschriebenen Metriken und Distanzmaße erlaubt und Anfragen in entsprechende Anweisungen der Abfragesprache des Datenbankmanagementsystems (z.B. SQL-Anfragen) übersetzt.

Um eine leichte Nutzbarkeit des Systems zu gewährleisten, kommt darüber hinaus eine GUI-Komponente zum Einsatz, die einen graphischen Zugang zum System anbietet und eine Visualisierung der Modelle und Anfrageergebnisse ermöglicht.

Hier wird davon ausgegangen, dass die Entwicklung von neuen Modellierungswerkzeugen für die semantische Attributierung ökonomisch nicht sinnvoll und weiterhin aus Gründen der Akzeptanz nicht zielführend ist. Stattdessen wird im Rahmen dieses Architekturvorschlags der Ansatz verfolgt, bestehende Modellierungswerkzeuge und CASE-Tools über eine Schnittstelle an das Informationsmodell-Datenbanksystem anzubinden. Die Anbindung erfolgt in Hinblick auf ein einfaches Schreiben von Modellen in die Modelldatenbank, ein Lesen von Modellen aus der Modelldatenbank und erweiterten Funktionen zur Attributierung von Informationsmodellen. Hierbei soll es über Systemerweiterungen möglich sein, innerhalb des Modellierungswerkzeugs eine Zuordnung von Modellelementen zu einzelnen Ontologiekonzepten vorzunehmen. Außerdem sollen die erweiterten Attribute eines Modellelements editierbar sein.

## 9 Anwendungsbereiche

### 9.1 Informationsmodell-basiertes Benchmarking

Als Benchmarking bezeichnet man den kontinuierlichen Prozess, bei dem Produkte, Dienstleistungen und insbesondere Prozesse und Methoden betrieblicher Funktionen über mehrere Unternehmen hinweg verglichen werden. Dabei sollen Unterschiede zu anderen Unternehmen offen gelegt, Ursachen für die Unterschiede und Möglichkeiten zur Verbesserung aufgezeigt sowie wettbewerbsorientierte Zielvorgaben ermittelt werden. Der Vergleich findet mit Unternehmen statt, welche die zu untersuchende Methode oder den Prozess hervorragend beherrschen.<sup>39</sup> Beim klassischen Benchmarking stehen folglich Metriken und die Abweichungen vom Branchenführenden im Vordergrund. Die Schwachstelle des Ansatzes besteht darin, dass nicht deutlich wird, *wie* es zu einer Realisierung von Verbesserungspotenzialen kommen kann, da zwar gemessen wird, wie effizient und effektiv ein Unternehmen bzw. Unternehmensteil agiert, aber nicht festgehalten wird, auf welche Weise die Effizienz- und Effektivitätsvorteile gewonnen werden. Es mangelt folglich an einem direkten Vergleich der zum Erfolg führenden Aufbau- und Ablaufstrukturen oder Prozessergebnissen. Aus diesem Grund wird vermehrt versucht, im Rahmen von Benchmarkingaktivitäten auch auf die Beschreibung von qualitativen Unterschieden zwischen den Benchmarkingobjekten abzustellen. Ein qualitativer Vergleich dieser Objekte ist jedoch sehr aufwändig und erfordert eine sehr gute Kenntnis des Benchmarkingobjektes.

---

<sup>39</sup> Vgl. Horváth, P., Herter, R.N. (1992), S. 5.

Der hier vorgeschlagene Ansatz des informationsmodell-basierten Benchmarkings beruht auf der Überlegung, dass Informationsmodelle eine gute Basis für das Benchmarking darstellen. Sofern diese um quantitative Erfolgs- und Leistungsgrößen angereichert werden (vgl. Abschnitt 6), kann auf Basis der semantischen Attributierung ein automatisiertes Benchmarking erfolgen. Hierzu werden systematisch Informationsmodelle (zumeist Prozessbeschreibungen) gesammelt, semantisch attribuiert und mit einheitlichen Erfolgs- und Leistungsgrößen angereichert. Bei diesen Erfolgs- und Leistungsgrößen kann es sich z.B. handeln um:

- Durchlauf- und Bearbeitungszeiten,
- Wartezeiten,
- Fehlerhäufigkeiten,
- Ressourcenbedarfe,
- Kosten oder
- Durchlaufwahrscheinlichkeiten.

Diese Modelle werden dann in einer Informationsmodelldatenbank hinterlegt und stehen dann für automatisierte Benchmarkinganalysen zur Verfügung. Unter Rückgriff auf die Informationsmodelldatenbank können z.B. für ein neues, semantisch attribuiertes Informationsmodell folgende Fragen beantwortet werden:

- Welche Informationsmodelle sind (syntaktisch und semantisch) ähnlich und welche Erfolgs- und Leistungsgrößen haben diese Modelle?
- Welche Informationsmodelle sind im Hinblick auf den Realweltausschnitt am erfolgreichsten und inwieweit unterscheiden sie sich vom neuen Modell?
- Welche nahe liegenden Optimierungspotenziale gibt es in Bezug auf vorherrschende, erfolgreiche Prozessmuster?

Bei der Beantwortung dieser Fragen ist es prinzipiell unerheblich, ob es sich um ein gesamtes komplexes Unternehmensmodell, einen Geschäftsprozess oder nur einen Prozessausschnitt handelt.

## **9.2 Erfahrungsdatenbanken**

Eine strukturell ähnliche Anwendung stellen der Aufbau und die Nutzung von Erfahrungsdatenbanken dar. Hierbei geht es darum, in Bezug auf ein konkretes Modellierungsziel Informationsmodelle in einer Informationsmodell-Datenbank zu identifizieren, die für die Erreichung dieses Ziels von Nutzen sein können. Im Vordergrund steht dabei der Lernaspekt des Subjektes. Dieses wird über die Auseinandersetzung mit den gefundenen Modellen in die Lage versetzt, ein individuelles, problemangemessenes Informationsmodell zu entwickeln. Das Subjekt nutzt hierzu das in den bestehenden Modellen codierte Wissen und vermeidet Fehler bei der Modellkonstruktion.

Die Verwendung von Erfahrungsdatenbanken für die Modellkonstruktion ist besonders dann von Vorteil, wenn kein geeignetes Referenz-Informationsmodell vorliegt. In diesem Fall kann die Beschäftigung mit Informationsmodellen, die ebenfalls eine Rekonstruktion des betreffenden Realweltausschnittes darstellen, eine Qualitätsverbesserung der Modellkonstruktion erzielt werden. Diese Verbesserung ergibt sich daraus, dass (1) plausibilisiert werden kann, dass keine wichtigen Modellbestandteile übersehen bzw. vergessen werden und (2) eine Identifikation besonders geeigneter Strukturmuster möglich wird.

### **9.3 Information Model Mining**

In Anlehnung an den Begriff des Data Mining wird hier der Begriff des Information Model Mining verwendet, um Verfahren zu bezeichnen, mit denen aus einer Menge von Informationsmodellen automatisch neue Erkenntnisse gewonnen werden können. Diese Erkenntnisse beziehen sich im Wesentlichen auf syntaktische und semantische Strukturmuster innerhalb der Modelle, können aber auch – eine entsprechende erweiterte Attributierung vorausgesetzt – weitergehende Charakteristika von Informationsmodellen sowie den zugrunde liegenden Realweltausschnitt betreffen. Hier sind insbesondere die oben im Kontext des informationsmodell-basierten Benchmarking erwähnten Metriken zu erwähnen.

Die mit dem Information Model Mining gewonnenen Erkenntnisse stellen statistische Aussagen zum Zusammenhang von Modellstrukturen und solchen Attributen dar. So kann beispielsweise mit dem Ansatz des Information Model Mining automatisch nach Strukturmustern gesucht werden, die als besonders effizient einzustufen sind oder besonders häufig verwendet werden.

### **9.4 (Halb-)Automatische Referenzmodellkonstruktion**

Bei der (halb-)automatischen Referenzmodellkonstruktion kann eine Informationsmodell-Datenbank eine sehr weitreichende Hilfe sein, indem auf Basis der mittels Datenbank dokumentierten, empirischen Fälle automatisch Vorschläge für Referenzmodellstrukturen generiert werden. Hierbei wird wie folgt vorgegangen:

1. Definition eines Realweltausschnittes, für den ein Referenzmodellvorschlag erzeugt werden soll. Diese Definition erfolgt anhand der Auswahl von Ontologiekonzepten.
2. Suche nach Informationsmodellen, die den spezifizierten Realweltausschnitt abdecken.
3. Identifikation dominanter Strukturmuster innerhalb dieser Informationsmodelle.
4. Montage dieser Strukturmuster zu einem Informationsmodell, d.h. einem Referenzmodellvorschlag.

Die automatisierte Konstruktion von Referenzmodellstrukturen darf nicht darüber hinwegtäuschen, dass eine manuelle Nachbearbeitung der Vorschläge durch den Modellierer erforderlich ist. Es ist außerordentlich unwahrscheinlich, dass automatisch generierte Strukturmuster bereits die notwendige Qualität aufweisen, die für ein Referenzmodell erforderlich ist. Darüber hinaus sind einzelne, für Ausschnitte einer Problemdomäne generierte Muster zu integrieren und dokumentieren.

## **9.5 Automatisierte Softwareauswahl**

Sofern in einer Informationsmodell-Datenbank die semantisch attribuierten Referenzmodelle von Standardsoftwaresystemen abgelegt werden, ist eine weitgehend automatische Software(vor)auswahl möglich. Hierzu definiert der Anwender diejenigen Bestandteile der Ontologie, die für ihn in Bezug auf den Soll-Funktionsumfang der Standardsoftware relevant sind. Danach ist es ein Leichtes festzustellen, welche Referenzmodelle diesen Soll-Funktionsumfang abdecken und wie weit diese Abdeckung reicht. Hierzu ist lediglich zu prüfen, wie groß die „semantische Schnittmenge“ des Soll-Funktionsumfangs der Software und des Referenzmodells ist.

## **10 Zusammenfassung und Ausblick**

In dieser Arbeit wurde gezeigt, wie durch eine zielgerichtete ontologie-basierte Attributierung von Informationsmodellen die Grundlage für eine automatisierte Verarbeitung von Informationsmodellen auf semantischer Ebene geschaffen werden kann. Aus dem dargestellten Konzept ergeben sich eine Reihe von innovativen Anwendungsbereichen, die bei einer herkömmlichen subjektgebundenen Interpretation von Informationsmodellen nicht realisierbar sind. Bei allen skizzierten Anwendungen ist zu berücksichtigen, dass ein nicht unerheblicher Aufwand für die Aufbereitung der Informationsmodelle betrieben werden muss. Diesem hohen, aber einmaligen Aufwand steht ein vergleichsweise hoher Nutzen mit nur geringen Kosten bei den folgenden Informationsmodellverwendungen gegenüber. Wirtschaftlich wird eine semantische Attributierung folglich insbesondere dann sein, wenn die aufbereiteten Informationsmodelle besonders häufig verwendet werden. Das ist jedoch nur in Szenarien zu erwarten, in denen eine superorganisationale Nutzung vorgesehen ist. Beispielhaft sei hier eine branchenbezogene Benchmarking-Datenbank angeführt, in der Informationsmodelle aus einer Vielzahl von Unternehmen abgelegt werden und die durch eine Clearing-Stelle betrieben wird.

Das Konzept der semantischen Attributierung von Informationsmodellen wirft eine Vielzahl von offenen Forschungsfragen auf, von denen die wichtigsten im Folgenden kurz zusammenfassend skizziert werden:

1. Der Begriff der Problemangemessenheit von Ontologien ist zu stellen und zu beantworten. Hier ist insbesondere zu untersuchen, über welche Eigenschaften eine Ontologie verfügen muss, damit sich ihre Verwendung im Rahmen der semantischen Attributierung von Informationsmodellen anbietet.
2. Die Qualität der Ergebnisse einer automatisierten Verarbeitung von Informationsmodellen wird maßgeblich von der Qualität der Zuordnung von Ontologiekonzepten zu den Informationsmodellbestandteilen bestimmt. Es ist zu klären, wie eine solche Qualität zu messen und sicherzustellen ist. Die im Rahmen dieser Arbeit eingeführten Integritätsbedingungen werden für reale Anwendungen voraussichtlich nicht ausreichen.
3. Es ist ein geeignetes Meta-Metamodell zu entwickeln, das für die semantische Attributierung von Informationsmodellen geeignet ist.
4. Es sind innovative Operationen und Algorithmen zu entwickeln, mit denen semantisch attributierte Informationsmodelle analysierbar sind. Hier spielen vor allem Metriken, Ähnlichkeits- und Distanzmaße eine wichtige Rolle.
5. Auf Basis der Operationen und Algorithmen ist eine Abfragesprache zu entwickeln, mit deren Hilfe die Inhalte von Informationsmodell-Datenbanken analysiert und ausgewertet werden können.
6. Die hier skizzierte Architektur einer allgemeinen Informationsmodell-Datenbank ist zu verfeinern, prototypisch umzusetzen und zu evaluieren.
7. Bestehende Modellierungswerkzeuge sind so zu erweitern, dass eine semantische Attributierung sowie der Import und Export von Informationsmodellen aus bzw. in die Informationsmodell-Datenbank möglich wird.
8. Es ist zu untersuchen, wann die semantische Attributierung von Informationsmodellen als wirtschaftlich zu bezeichnen ist.
9. Es sind geeignete formale ontologie-basierte Sprachen zur Repräsentation von entsprechender Domänenontologien zu entwickeln bzw. weiterzuentwickeln.
10. Es sind geeignete Benutzerschnittstellen für neue Abfragemechanismen und für die Präsentation von Such- bzw. Klassifikationsergebnissen zu entwickeln.

## Literatur

Bergamaschi, S., Castano, S., Beneventano, D., Vincini, M. (2001): Semantic integration of heterogeneous information sources, Special Issue on Intelligent Information Integration, Data & Knowledge Engineering, 2001, Vol. 36, No. 1, S. 215-249.

Bisson, G. (1995): Why and how to define a similarity measure for object based representation systems. Towards Very Large Knowledge Bases, 1995.

Chidamber, S.R., Kemerer, C.F. (1994): A Metrics Suite for Object Oriented Design, IEEE Trans. Software Eng. 20 (6) 1994, S. 476-493.

Chidamber, S.R., Darcy, D.P., Kemerer, C.F. (1998): Managerial Use of Metrics for Object-Oriented Software: An Exploratory Analysis, IEEE Trans. on Software Eng. 24 (8) 1998, S. 629-639.

De Bruijn, J., Ding, Y., Arroyo, S., Fensel, D. (2004): Semantic Information Integration in the COG Project, DERI Technical Report 2004-05-04, University of Innsbruck 2004.

Ehrig, M., Sure, Y., Staab, S. (2005): Supervised Learning of an Ontology Alignment Process, in: Althoff, K.-D., Dengel, A., Bergmann, R., Nick, M., Roth-Berghofer, T. (Eds.): Professional Knowledge Management: Third Biennial Conference, WM 2005, Kaiserslautern, Germany, April 10-13, 2005, Revised Selected Papers, S. 508-517.

Electronic Industries Association (Eds.) (1994): CDIF – CASE Data Interchange Format; verfügbar unter <http://www.eigroup.org>.

Eriksson, H., Puerta, A., Musen, M. (1994): Generation of Knowledge Acquisition Tools from Domain Ontologies, Proceedings of the 8th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop; Banff, Canada 1994.

Gennari, J., Tu, S., Rothenfluth, T., Musen, M. (1994): Mapping Domains to Methods in Support of Reuse, International Journal of Human and Computer Studies, o.Jg. (1994) 41, S. 399 - 424.

Fellbaum, C. (1998): WordNet: An Electronic Lexical Database (Language, Speech, and Communication); Cambridge, MA: The MIT Press, 1998.

Fensel, D., Groenboom, R. (1997): Specifying Knowledge-based Systems with Reusable Components, Proceedings 9th Int. Conference of Software Engineering and Knowledge Engineering (SEKE '97) Madrid 1997.

- Gruber, T.R. (1993): A translation approach to portable ontology specifications, Knowledge Acquisition, 5, S. 199-220.
- Gruber, T.R. (1995): Toward principles for the design of ontologies used for knowledge sharing, International Journal of Human-Computer Studies 43 (1995) 5/6, S. 907-928.
- Heinrich, L.J. (2001): Wirtschaftsinformatik, Einführung und Grundlegung, 2. vollst. überarb. und erg., München, Wien: Oldenbourg, 2001.
- Horváth, P., Herter, R. N. (1992): Benchmarking – Vergleich mit den Besten der Besten, Controlling, Jg. 4, 1992, Nr. 1, S. 4-11.
- Lehner, F. (1995): Grundfragen und Positionierung der Wirtschaftsinformatik, in: Lehner, F., Hildebrand, K., Maier, R. (Hrsg.), Wirtschaftsinformatik. Theoretische Grundlagen, München, Wien: Hanser, 1995, S. 1-72.
- Lorenz, M., Kid, J. (1994): Object Oriented Software Metrics, Prentice Hall Publishing, 1994.
- Noy, N.F., Musen, M.A. (1999): Smart: Automated support for ontology merging and alignment, Technical Report SMI-1999-0813, Stanford Medical Informatics, 1999.
- Object Management Group (1999): OMG Unified Modeling Language Specification; verfügbar unter <ftp.omg.org/pub/docs/ad/99-06-09.pdf>.
- Object Management Group (2003): Meta Object Facility (MOF), Version 1.4., März 2003.
- Preece, A.D., Hui, Kit ying., Gray, W.A., Marti, P., BenchCapon, T.J.M., Cui, Z. Jones, D. (2001): Kraft: An Agent architecture for knowledge fusion, International Journal of Cooperative Information Systems, Vol. 10, 2001, No. 1-2, S. 171-195.
- Reeves, B. N., Hayashi, Y. (1999): Looking Good: Exploring Visual Representations of Object Oriented Metrics, SoftVis99: 3<sup>rd</sup> Workshop on Software Visualization, University of Technology, Sydney, Australia, Dec. 1999.
- Rosenberg, L. H. (1998): Applying and Interpreting Object Oriented Metrics, Software Assurance Technology Center, 1998.
- Scherer, G. (1996): Einführung in die Philosophie; Düsseldorf u.a. 1996.
- Seedorf, S., Tomczyk, P. (2005): Ein angepasstes Vorgehensmodell zur Entwicklung von Domänenontologien in der semantischen Softwareentwicklung, Arbeitspapier 2/2005 (März 2005), Arbeitspapiere in der Wirtschaftsinformatik, Lehrstuhl für Wirtschaftsinformatik III, Universität Mannheim und FZI Forschungszentrum Informatik, Forschungsbereich Datenbanksysteme, Karlsruhe, 2005, verfügbar unter: [http://www.collabawue.de/cms/dokumente/Arbeitspapier\\_Seedorf\\_Tomczyk\\_SBSE.pdf](http://www.collabawue.de/cms/dokumente/Arbeitspapier_Seedorf_Tomczyk_SBSE.pdf)

Schütte, R. (1998): Grundsätze ordnungsmäßiger Referenzmodellierung, Konstruktion konfigurations- und anpassungsorientierter Modelle; Wiesbaden: Gabler, 1998.

Strahringer, S. (1998): Ein sprachbasierter Metamodellbegriff und seine Verallgemeinerung durch das Konzept des Metaisierungsprinzips, in: Pohl, K., Schürr, A., Vossen, G. (Hrsg.), Proceedings des GI-Workshops Modellierung '98, 1998.

Studer, R., Fensel, D., Decker, S., Benjamins, V.R. (1999): Knowledge Engineering – Survey and Future Directions, in: Puppe, F. (Eds.): XPS-99: Knowledge-Based Systems – Survey and Future Directions, Proceedings of the 5<sup>th</sup> Biannual German Conference on Knowledge-Based Systems, Berlin et al. 1999, S. 1-23.

van Elsuwe, H., Schmedding, D. (2003): Metriken für UML-Modelle, Informatik Forschung und Entwicklung (2003) 18, S. 22-31.

Wielinga, B.J., Schreiber, T.A. (1994): Conceptual Modelling of Large Reusable Knowledge Bases, in: Luck, K.V., Marburger, H. (Eds.): Management and Processing of Complex Data Structures, Lecture Notes in Computer Science, Bd. 777; Berlin u.a. 1994, S. 181-200.

Yao, H., Orme, A. M., Etzkorn, L. (2005): Cohesion Metrics for Ontology Design and Application, Journal of Computer Science 1 (1) 2005, S. 107-113.

Zimmermann, H.-J., Angstenberger, J., Detro, F., Ivens, M., Kasper, C., Liever, K., Meier, W., Minas, H.-J., Teitscheid, P., Weber, R. (Hrsg.): Datenanalyse: Anwendung von DataEngine mit Fuzzy-Technologien und Neuronalen Netzen; Düsseldorf 1995.

**Folgende Arbeitsberichte des ISPRI sind bisher erschienen:**

**Arbeitsbericht 1:** Ahlemann, F.; Teuteberg, F.; Schroeder, C.: Kompetenz- und Reifegradmodelle für das Projektmanagement: Grundlagen, Vergleich und Einsatz; in: Ahlemann, F., Teuteberg, F. (Hrsg.): ISPRI-Arbeitsbericht Nr. 01/2005 des Forschungszentrums für Informationssysteme in Projekt- und Innovationsnetzwerken (ISPRI), März 2005. (ISBN: 3-936475-24-5)

**Arbeitsbericht 2:** Ahlemann, F.; Teuteberg, F.; Brune, G.: Ontologie-basierte Attributierung von Informationsmodellen: Grundlagen und Anwendungsgebiete; in: Teuteberg, F.; Ahlemann, F. (Hrsg.): ISPRI-Arbeitsbericht Nr. 01/2006 des Forschungszentrums für Informationssysteme in Projekt- und Innovationsnetzwerken (ISPRI), Mai 2006. (ISBN: 3-936475-26-1)