

Dynamische Typisierung und non-monotonic Reasoning: Voraussetzungen für semantische Interoperabilität

Straub HR¹, Lehmann M¹

¹Semfinder AG, Kreuzlingen

Abstrakt

Semantische Interoperabilität ist für eHealth eine Voraussetzung. Dabei spielen nicht nur die verschiedenen Datenformate eine Rolle, sondern die Semantik der Dokumente selber, d.h. ihr Bedeutungsinhalt, muss interoperabel sein. Die vorliegende Arbeit erläutert die prinzipiellen Probleme, die mit einem Standard einhergehen. Als Alternative wird ein dynamischer Ansatz vorgestellt, der auf Standards verzichtet, bzw. auch in einer Umgebung mit heterogenen Standards funktioniert. Die Basis eines solchen dynamischen Ansatzes ist eine komposite Begriffsdarstellung auf der Basis einer multidimensionalen und multifokalen Begriffsarchitektur. Als Schlussmechanismus wird eine nicht-monotone Logik mit einer Chronologiekontrolle über ein dynamisches Triggerbegriffssystem eingesetzt.

Abstract

Semantic interoperability is a crucial condition for eHealth. Not only formal aspects of the documents have to be considered, but the semantics, e.g. the meaning of the mentioned facts, must be made interoperable. We show the main problems, which arise when a static standard is chosen as a medium for interoperability. An alternative approach is introduced, not a standard, but a method to interoperate between given heterogeneous standards. The approach is based on a multidimensional and multifocal concept architecture. Concepts are presented in a strict composite form, in order to deal with the multidimensional nature of semantics. We explain why dynamic type-definitions are essential for the given task and why the reasoning must be non-monotonic. When operating on an input, the non-monotonic reasoning system is controlled by a system of few dynamic trigger concepts.

Problem: Semantischer Standard

Proprietäre Formate von KIS- und Praxissoftware behindern den Datenaustausch erheblich. Dazu kommen unterschiedliche Sprachen und unterschiedliche Anforderungen an die Dokumentation. Längst ist erkannt, dass diese Probleme nicht durch eine Überbrückung nur der formalen Standards gelöst werden können, viel schwerer wiegt die Tatsache, dass die Inhalte und Bedeutungen der dokumentierten Daten weder standardisiert noch formal geregelt sind. Unter dem Schlagwort "semantische Interoperabilität" [1,2,3,4,5] summieren die Versuche, hier Abhilfe zu schaffen.

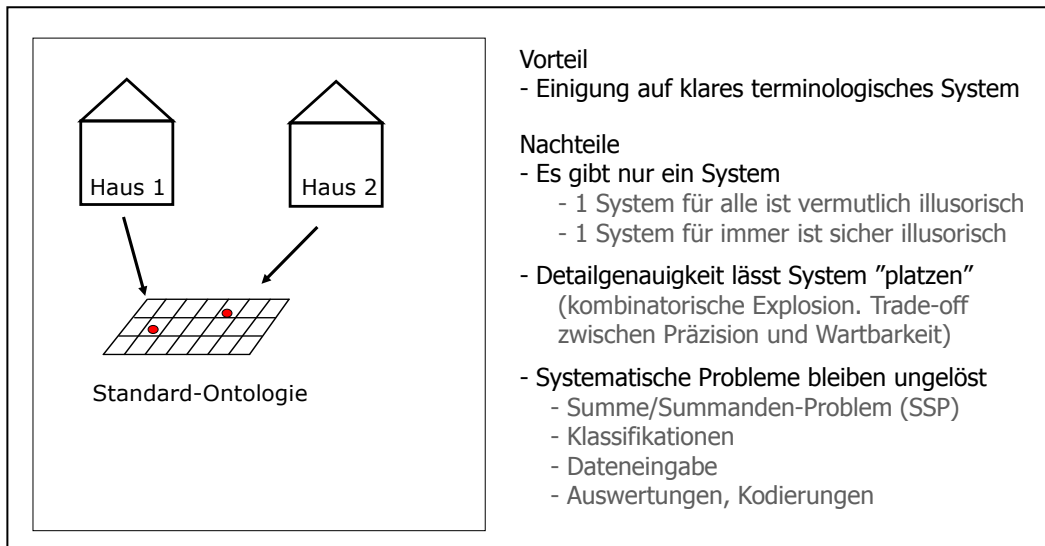


Abb. 1.: Statischer Ansatz: Datenaustausch über Abbildungen in einen "semantischen" Standard

Bei näherem Hinsehen wird klar, dass das Problem oft nur verschoben wird. So wird meistens ein Standard genannt, der die Semantik vereinheitlichen soll (siehe Abb. 1.), je nach Aufgabe ist dies eine Klassifikation (ICD-10), eine aufgabenorientierte Norm (LOINC [6]) oder eine standardisierte Terminologie (SNOMED CT [7]). Doch können diese Standards die gestellte Aufgabe lösen?

Das Problem der Standards liegt an mehreren Stellen:

- A) Kontinuierlicher Anpassungsbedarf: Obwohl ein Standard naturgemäß unveränderlich ist und bleiben muss, ist in der Praxis sofort ein Erweiterungsbedarf und Veränderungsbedarf festzustellen. Selbst ein so ausgebauter Standard wie SNOMED CT kann nicht alle Bedürfnisse befriedigen.
- B) Versionswildwuchs: Dieser folgt unmittelbar und zwingend aus dem eben unter A) genannten Anpassungsbedarf, sowie aus den je nach Zielgruppe unterschiedlichen Fragestellungen. Die ICD-10 Versionen DE, CH, USA, UK und AT unterscheiden sich z.T. massiv. Deutschland gibt jedes Jahr eine oft radikal erneuerte ICD-10-Version heraus. Eine Interoperabilität über unterschiedliche Versionen ist jedoch per se unpräzise.
- C) Mangelhafte Struktur: Eine nähere Beschäftigung mit Semantik zeigt, wie entscheidend die Begriffsarchitektur ist. Um einen Begriffsbereich wie die medizinischen Diagnosen darzustellen, ist mindestens eine multifokale Architektur notwendig. Eine bloße Aufzählung (Thesaurus) oder eine Hierarchie (ICD-10) führen schnell zu gravierenden Interpretationsproblemen. Die bessere Alternative, SNOMED CT, enthüllt im Kern eine multidimensionale Architektur, was für eine semantische Modellierung ebenfalls nicht ausreicht.
- D) Klassifikationsbedarf: Information muss für Vergleiche immer gruppiert werden. Entgegen einer weitverbreiteten Meinung kann aus einer feingranulären Information aber nicht ohne zusätzliche Regeln auf eine grobgranuläre geschlossen werden (Unterschiedliche Perspektiven, Summe-Summanden-Problem[8]). Je feingranulärer der Standard ist, desto weiter entfernt von den gesuchten "Schubladen" der grobgranulären Klassifikation ist er. Je grobgranulärer er andererseits ist, desto wahr-

scheinlicher wird es, dass der gewählte allgemeine Standard dem gerade aktuellen Klassifikationsbedürfnis nicht entspricht. Denn Information kann immer unterschiedlich gruppiert werden. Darin liegt auch der Grund, dass Hierarchien für medizinisch-ärztliche Informationen nicht logisch widerspruchsfrei konstruiert werden können (siehe Punkt C). Ohne Willkür kann aus feingranulärer Information nicht auf grobgranuläre geschlossen werden. Deshalb erfüllen feingranuläre Terminologien wie SNOMED CT den Klassifikationsbedarf nicht aus sich selber heraus, sondern brauchen zusätzliche Regeln [8,9], die über ein konventionelles Mapping deutlich herausgehen.

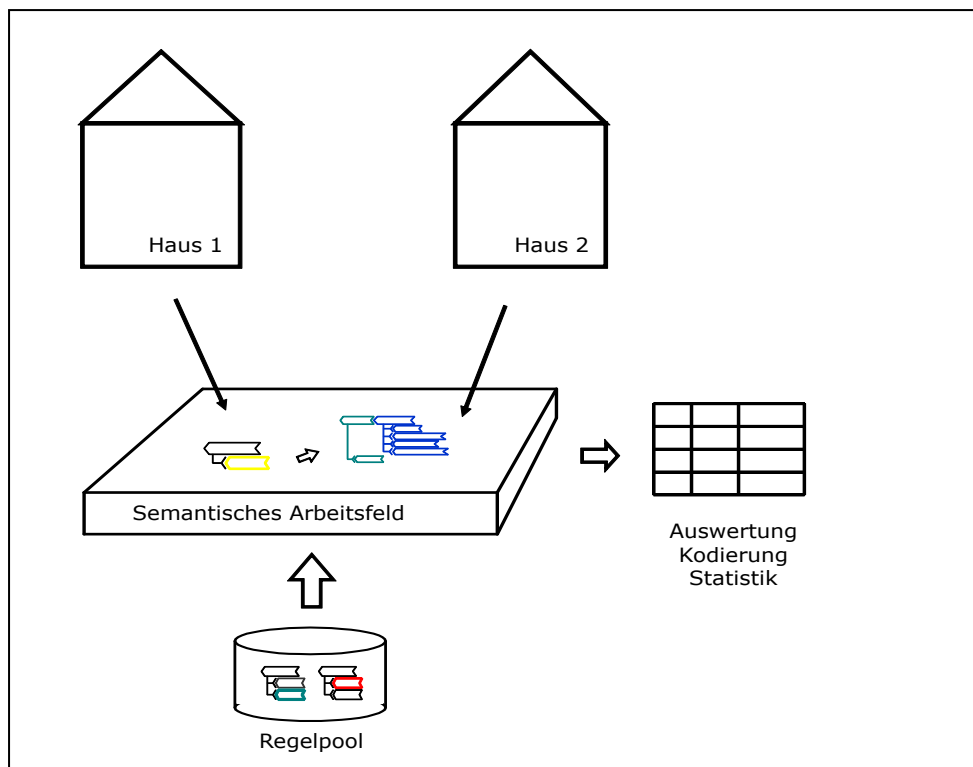


Abb. 2.: Dynamischer Ansatz: Kein vordefiniertes Gitter, sondern Regelset zur Umformung

Dynamischer Ansatz

Die oben genannten Problem lassen sich durch einen dynamischen Ansatz (Abb. 2) lösen. Dabei wird akzeptiert, dass die interoperabel zu verbindenden Quellen unterschiedliche semantische Konzeptionen verwenden. Die Aufgabe der Interoperabilität liegt nun darin, die unterschiedlichen Ansätze jeweils vollständig (!) zu beschreiben und sie an genau den Stellen ineinander überzuführen, wo dies möglich ist. Wo eine Überführung in einer feingranulären Form nicht möglich ist, wird für den Informationsaustausch ein Abgleich auf dem nächst grobgranuläreren Niveau gesucht. Dabei müssen natürlich unterschiedliche Perspektiven beachtet werden. Das der Interaktion zugrunde liegende Datenmodell muss genuin komposit (postkoordiniert), multidimensional und multifokal [10,11] sein, sonst ist die Transformation nicht möglich. Dass der dynamische Ansatz die Aufgabe der semantischen Interoperabilität wirklich lösen kann, beweist das erfolgreiche One-Step-Coding[12] der ICD-10. Denn bei dieser Aufgabe müssen die unterschiedlichen semantischen Ansätze, die hinter der freien Textformulierung der sehr unterschiedlichen Ärzte, Fachrichtungen und Schulen stecken, und die unterschiedlichen Klassifikationsanforde-

rungen verschiedener ICD-10-Varianten erkannt und ineinander überführt werden – ein genuin dynamisches Problem der Interoperabilität, das keinesfalls durch einen neuen Standard, sondern nur durch eine Methode der kontrollierten Umformung gelöst werden kann. Im folgenden werden zwei Charakteristika aufgeführt, die eine solche dynamische Lösung des Problems im Unterschied zur statischen Standard-basierten auszeichnen: die dynamische Typisierung und das nicht-monotone Reasoning.

Dynamische Typisierung

Kann ich aus Finger auf Hand schliessen? Eine Lokalisation am Finger ist eine Lokalisation an der Hand. Eine Amputation eines Fingers ist aber nicht eine Amputation einer Hand. Der Unterschied liegt in der Typisierung: Der Begriff *Finger* ist – wie prinzipiell alle Begriffe – keinesfalls ein eindeutiger semantischer Punkt. Er kann unterschiedliche Oberbegriffe (Typen) annehmen, und dadurch unterschiedliche Semantiken repräsentieren. Ist der Typ *Lokalisation*, darf von der feingranulären Lokalisation (Finger) auf die grobgranuläre (Hand) geschlossen werden. Beim Typ *Amputationsobjekt* ist dies aber nicht der Fall, vom Amputationsobjekt *Finger* darf nicht auf das Amputationsobjekt *Hand* geschlossen werden. Das Beispiel macht klar, dass Semantik nur funktioniert, wenn die Begriffe klar typisiert werden und dass diese Typisierung vom Kontext abhängt, d.h. prinzipiell variabel sein muss.

Die semantische Interoperabilität ist ganz besonders auf eine dynamische Typisierung angewiesen, da Inhalte aus heterogenen Quellen und somit potentiell unterschiedliche Sichtweisen verglichen werden. Fixe Standards sind nicht in der Lage, dynamisch zu typisieren. Dies trifft auf alle Standards mit hierarchischem Grundaufbau zu, sowie auf solche mit fixen Achsen (Dimensionen), also neben Klassifikation wie ICD-10 auch auf wesentlich entwickeltere wie SNOMED-CT. Das Problem ist ein grundlegendes und betrifft jeden Standard. Ein solcher muss prinzipiell jeden seiner Begriffe klar und für immer einordnen. Für immer einordnen heisst, fix zu typisieren. Für Interoperabilität und für Auswertungen nach unterschiedlichen Gesichtspunkten und in unterschiedlichen Kontexten ist aber eine dynamische Typisierung eine notwendige Bedingung.

Non-Monotonic Reasoning (NMR)

Eine dynamische Typisierung funktioniert nur mit einer nicht-monotonen Logik. In der klassischen Logik sind Aussagen entweder wahr oder falsch und eine Änderung der Wahrheitszuordnung ist nicht möglich (monotone Logik). So ist z.B. die Prädikatenlogik der ersten Stufe (First order logic) streng monoton, dasselbe gilt für Subsets dieser Logik, wie z.B. für die Description logic (DL) nach Baader [13], die in Ontologien und für SNOMED CT verwendet wird. Umformungen, wie sie für die Überführung von einem Standard in einen anderen nötig sind und variable, kontextbezogene Interpretation verlangen aber nach einer Möglichkeit, Aussagen je nach Kontext unterschiedlich zu interpretieren. Eine solche Möglichkeit erlauben nicht-monotone Logiken. Eine Unterform davon sind Default-Logiken, dabei wird zuerst eine Defaultzuordnung gemacht (alle Vögel fliegen), die dann je nach Kontext überschrieben werden kann (Strausse fliegen nicht). Default-Logiken bieten, wenn sie zweckdienlich gestaltet werden, neben der besseren Modellierbarkeit auch Vorteile bei der Wartung und die Chance einer besseren Performance auf Computern. Allerdings sind nicht-monotone Logiken nicht so geschlossen wie monotone. Sie sind gegenüber der Zeit nicht invariant. Der Ablauf der Schlussfolgerung muss deshalb klar kontrolliert werden können (Chronologiekontrolle).

Beispiel eines NMR mit Chronologiekontrolle

Diese Kontrolle kann zum Beispiel über Trigger erfolgen. Für die automatisierte Kodierung verwenden wir seit 10 Jahren ein nicht-monotones Reasoning [12]. Die Eingabewörter des Arztes werden dabei sukzessive interpretiert. Sie werden als erstes typisiert (überschreibbar) und dann zu Begriffsclustern (Molekülen) miteinander verbunden (komposite, multidimensionale und multifokale Architektur), wobei nicht genannte, aber implizite Begriffe bestimmt und in das komposite Gewebe integriert werden (Siehe Abb. 3 und 4). Jeder Begriff im momentanen Interpretations-Zwischenstand kann einen **Trigger** darstellen, der zum Beispiel als Ausnahmekontext eine defaultmässige Annahme überschreibt. Der ganze Ablauf wird aber vor allem dadurch übersichtlich und kontrollierbar, dass einige wenige entscheidende Schlüsselbegriffe die Meilensteine darstellen, die bei Erreichen eines bestimmten Ziels entstehen und damit als Trigger ein neues Bündel von Regeln aktivieren. Die Entstehung der Trigger wird aus den jeweiligen Interpretationsstatus während der Inferenz dynamisch gesteuert. Als primäre Trigger werden dem Input bei Beginn die Ziele der Kodierung (z.B. ICD-10 in der Variante CH-2009) mitgegeben, die Interpretation des Textes erfolgt so in Richtung der entsprechenden Kodierung. Das Erfüllen des hohen Qualitätsanspruchs, der naturgemäss von der Anwenderseite her an ein One-Step-Coding gestellt wird, beweist, dass ein NMR mit dieser Art Chronologiekontrolle in einem semantisch sehr komplexen Umfeld erfolgreich sein kann.

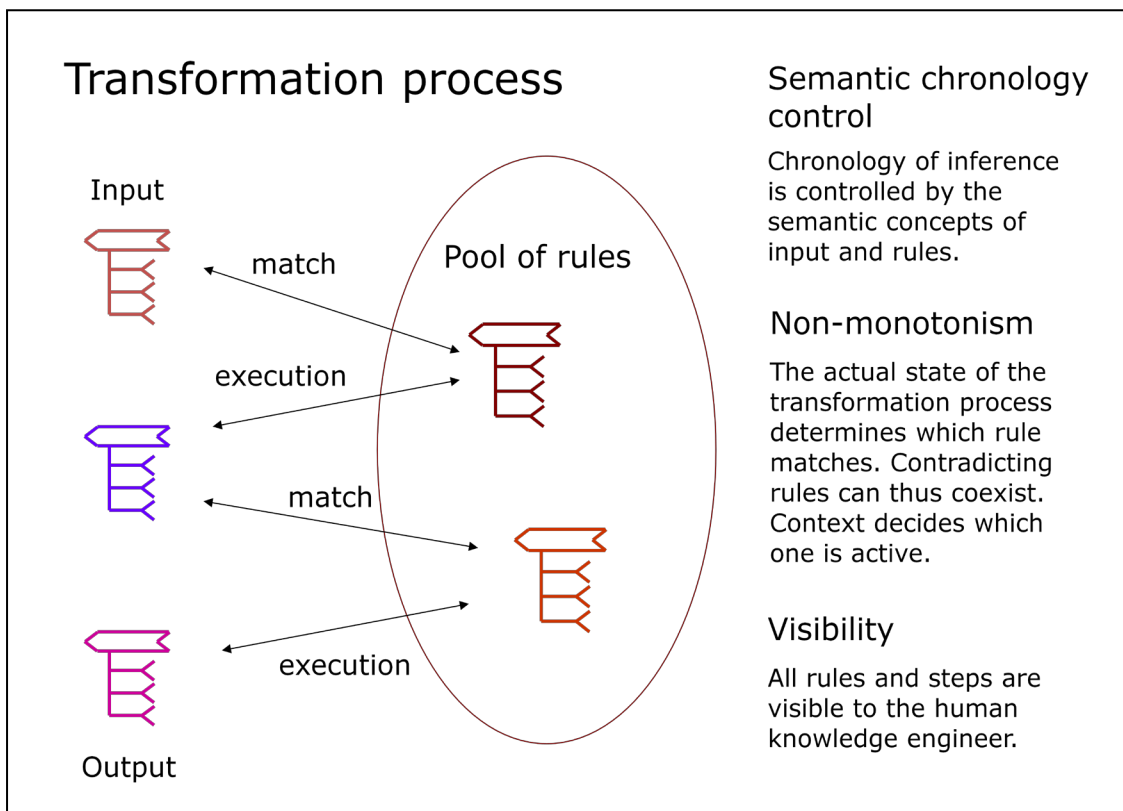


Abb.3.: Inferenzablauf von Input zu Output, mit Regelpool

Fraktur und Radius werden typisiert
und mit Attributen verbunden

Semantische Kerne
werden verbunden

Implizite
Information
wird dargestellt

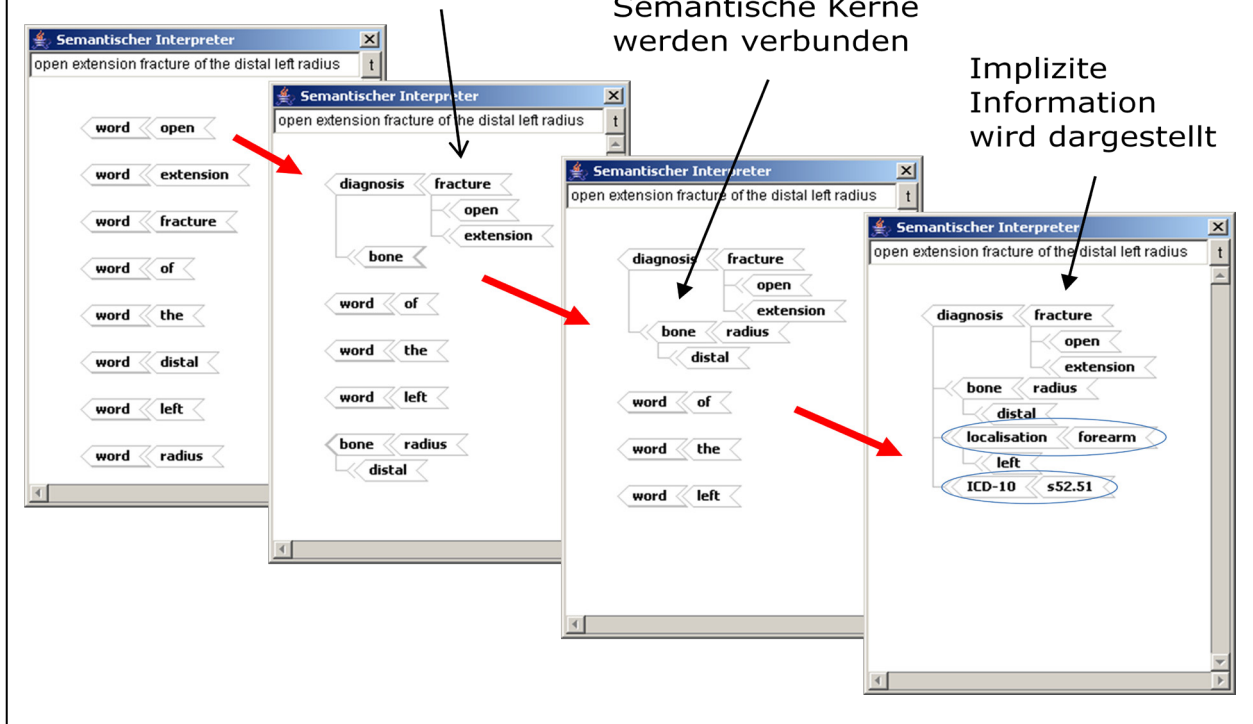


Abb.4.: Inferenzablauf: von den Wörtern zur Semantik

Fazit

Semantische Interoperabilität ist nur mit dynamischer Typisierung und NMR erreichbar. Description Logic und darauf basierende Verfahren sind zwar geeignet, die Konsistenz eines Begriffssystems zu testen, ihnen fehlt jedoch die notwendige Dynamik für komplexe Interpretationsaufgaben. Um ein Non-Monotonic-Reasoning System zu kontrollieren, kann eine Steuerung über dynamische Trigger eingesetzt werden. Diese Lösung bewährt sich im Praxiseinsatz.

Literatur

- [1] Rinner C (2010) Intramuraler Datenaustausch mittels ELGA-basierter CDA-Dokumente am Beispiel des AKH Wien, in: Schreier G, Hayn D, Ammenwerth E (Hrsg). Tagungsband der eHealth2010, 6.-7. Mai 2010, Wien.
- [2] Oemig F, Blobel B (2009) Semantic Interoperability between Health Communication Standards through Formal Ontologies, in: Adlassnig K-P, Blobel B, Mantas J, Masic I (Editors), Proceedings of the MIE 2009 - European Federation for Medical Informatics, IOS Press, ISBN: 978-1-60750-044-5, p.200-204.
- [3] Fraunhofer-Institut ISST (2010) White-Paper-Reihe "Semantische Interoperabilität": <http://www.isst.fraunhofer.de/geschaeftsfelder/gfegovernment/referenzprojekte/sie/white-paper/index.jsp> [Zuletzt besucht am 14. Mai 2010]
- [4] Blobel B, Pharow P (2009) Analysis and Evaluation of EHR Approaches, Methods of Information in Medicine 2009 48: 162-169
- [5] GMDS-Arbeitsgruppe "Standards für Interoperabilität und elektronische Gesundheitsakten", Workshop vom 19.4. 2010: <http://agsie.eh-cc.de/veranstaltungen/workshop-19-4-2010/> [Zuletzt besucht am 14. Mai 2010]
- [6] <http://loinc.org/> [Zuletzt besucht am 14. Mai 2010]
- [7] <http://www.ihtsdo.org/> [Zuletzt besucht am 14. Mai 2010]
- [8] Straub HR, Duelli M, Mosimann H, et al. (2006) From Terminologies to Classifications – the Challenge of Information Reduction, in: Proceedings of the European Federation for Medical Informatics Special Topic Conference, Timisoara, Romania, 2006: pp 341-352. See also: http://www.semfinder.com/fileadmin/Daten/Dateien/Publikationen/From_terminologies_to_classification.pdf [Zuletzt besucht am 14. Mai 2010]
- [9] Straub HR (2006) With Semantic Analysis from Noun Phrases to SNOMED CT and Classification Codes, an: Semantic Mining Conference on SNOMED CT in Copenhagen, Okt. 2006. http://www.semfinder.com/fileadmin/Daten/Dateien/Publikationen/Semantic_Analysis.pdf
- [10] Straub HR (2002) Four Different Types of Classification Models, in: Grütter R. (Hrsg.) Knowledge Media in Health Care: Opportunities and Challenges. Hershey / London: Idea Group Publishing, S. 58 – 82. ISBN: 1-930708-13-0
- [11] Straub HR (1998, 2001) Vier Basisarchitekturen für Klassifikationsmodelle: http://www.semfinder.com/fileadmin/Daten/Dateien/Publikationen/Vier_Basisarchitekturen.pdf [Zuletzt besucht am 14. Mai 2010]
- [12] <http://www.semfinder.com> [Zuletzt besucht am 14. Mai 2010]
- [13] Baader F et al (2003) The Description Logic Handbook, Theory, Implementation and Applications, Cambridge Press, ISBN: 0 521 78176 0

Korrespondenz

Hans Rudolf Straub
Semfinder AG
Hauptstrasse 53
CH-8280 Kreuzlingen
E-Mail: straub@semfinder.com