The Semantics of Procedures and Diseases in SNOMED® CT

S. Schulz[1], S. Hanser[1], U. Hahn[2], J. Rogers[3]
[1]Department of Medical Informatics, Freiburg University, Freiburg, Germany
[2]Jena University Language & Information Engineering (JULIE) Lab, Jena, Germany
[3]School of Computer Science, Manchester University, Manchester, United Kingdom

Summary

Objectives: SNOMED® CT is emerging as a reference terminology for the entire health care process. It claims to be founded on logic-based modelling principles. In this article, we analyze a special encoding scheme for diseases and procedures in SNOMED® CT, the so-called relationship groups (RGs), which had been devised to avoid ambiguities in definitions.
Methods: We reformulate SNOMED® CT's relationship groups in the format of description logics in order to check whether RGs serve the needs they were designed for.
Results: We show that a considerable proportion of relationship groups represent hidden mereological relations. We also report discrepancies encountered between the defined semantics of many SNOMED® CT terms and their intuitive meaning, as well as inconsistencies detected between the definition of various complex composed terms and the definition of their top-level parents.
Conclusions: We formulate recommendations for improving SNOMED® CT by replacing most occurrences of relation groups by formally more adequate "part-of" relations.

Keywords

SNOMED CT, knowledge representation, logic

Methods Inf Med 2006; 45: 354–8

1. Introduction

SNOMED® Clinical Terms (SNOMED® CT) is a huge clinical terminology constructed by merging, expanding, and restructuring the previous SNOMED version RT and the Clinical Terms Version 3 (former Read Codes). In the UK version (January 2006), SNOMED® CT contains 278,754 current concepts, 705,770 current terms, and 1,361,292 current relationships between concepts. In the coming years it will be deployed for routine usage in several countries (US, UK, Denmark), and it is intensively being analyzed by medical terminologists and decision makers in many other countries.

SNOMED® CT is a concept-oriented controlled vocabulary which has been designed according to previously suggested criteria for computer-based medical terminologies [1]. Its developers describe it as a clinical reference terminology, i.e. "a set of concepts and relationships that provides a common reference point for comparison and aggregation of data about the entire health care process" [2]. SNOMED® CT concepts belong to multiple taxonomic hierarchies and are related with one another by various semantic relationships such as *isA, hasAssociatedTopography, hasAction, hasAssociatedMorphology.* Interestingly, SNOMED® CT explicitly encodes "part-of" relationships only in its anatomy branch [4]. There is no relation with which to describe partonomy between diseases or procedures.

2. Relationship Groups in SNOMED®

SNOMED® RT and CT follow a formal semantics based on KRSS, an early description logic [5], and have used a terminological classifier for terminology development [6, 7]. In the current distributions (SNOMED, CLUE), however, the relational format of the MRREL table of the UMLS [8] requires SNOMED® CT content to be compressed into the common OAV (object-attribute-value triplets) format. For a logic-based model such a representation is ambiguous because it i) obscures which attribute-value pairs are *sufficient* for the definitions of semantic categories[a], ii) lacks role quantifications, and iii) does not indicate whether sets of related attribute-value pairs should be interpreted as disjunction, conjunction or optional. Table 1 gives an example of the resulting ambiguities in the OAV format.

Since 2004, the UMLS Rich Release Format (RRF) has encoded further information in the MRSAT table, including which semantic categories are defined or primitive, which OAV triplets are optional (qualifiers) and which are restrictions, and which relationship group each belongs to.

The purpose of relationship groups (also called "role groups" [6]) such as introduced

[a] Many ontological assumptions of SNOMED® CT are still unclear. For instance, different things like "*Foot*", "*Absent Foot*", "*Football (qualifier value)*", "*Europe*", "*Love*", "*mmol*", "*Yin excess*", "*Kiel Classification*", are "concepts" in SNOMED® CT. Not all of them are universals in a philosophical sense, or correspond to concepts or classes as used in description logics. Hence we use – for the sake of neutrality – the term "semantic category" for what SNOMED names "concept" [3].

by SNOMED® CT is best explained by an example: *RemovalOfForeignBodyFromTheStomachByIncision* involves *ForeignBody* as the value for the *hasDirectMorphology* attribute, *StomachStructure* as value of the attribute *hasProcedureSite*, and both *Incision* and *Removal* as values of the attribute *hasMethod*. However, such a simplistic representation is ambiguous: it could also be interpreted as *RemovalOfStomach* or *IncisionOfForeignBody*. In order to prevent such misleading interpretations, SNOMED® CT introduces so called "relationship groups" which declare associations between sets of OAV triplets (see Table 2). Relationship groups are identified by an integer value; this however does not imply any temporal or other ordering between groups.

The *SNOMED® CT Technical Implementation and Technical Reference* makes the following statement about relationship groups:

"Relationships, for a concept that are logically associated with each other. The Relationship group field in the Relationships Table is used to group these rows together for a concept."

As relationship groups occur in about 17,000 disease and 13,000 procedure descriptions, this phenomenon constitutes a major issue in SNOMED® CT. In [6], Spackman et al. propose a description logics representation for relationship groups, in which they are expressed by an anonymous relation, named *rg*. From an ontological point of view, the proposed solution is, however, rather obscure. In the following we therefore explore the possible semantics of SNOMED® CT relationship groups. We show that some basic assumptions of SNOMED® CT are ontologically problematic, and we propose a solution for clarification which will be mostly compatible with the current SNOMED® CT architecture.

Table 1 Example of OAV (object-tribute-value) format of SNOMED® CT such as distributed via the UMLS Metathesaurus

SNOMED® CT Concept 1	SNOMED® CT Relationship	SNOMED® CT Concept 2
RenalGlomerularDisease	*hasFindingSite*	*Kidney*
RenalGlomerularDisease	*hasOnset*	*GradualOnset*
RenalGlomerularDisease	*hasOnset*	*SuddenOnset*

Table 2 Entries in the January 2006 SNOMED® CT (UK release) core relationships table for "*RemovalOfForeignBodyFromTheStomachByIncision*", using three relationship groups

SNOMED® CT Concept 1	SNOMED® CT Relationship	SNOMED® CT Concept 2	RG
RemovalOfForeignBodyFromTheStomachByIncision	*access*	*OpenApproach*	*0*
	isA	*RemovalOfForeignBodyFromStomach*	*0*
	isA	*IncisionOfStomach*	*0*
	method	*RemovalAction*	*1*
	directMorphology	*ForeignBody*	*1*
	procedureSite	*StomachStructure*	*1*
	method	*IncisionAction*	*2*
	procedureSite	*StomachStructure*	*2*

3. Ontological Analysis of Relationship Groups

We refer to the same parsimonious variant of description logics as used by Spackman et al. [6]. Names of semantic categories are characterized by initial capital letters. They can be joined by the AND operator. As an example, the expression *AcuteDigestiveSystemDisorder AND AcuteInflammatoryDisease* denotes inflammatory diseases of the digestive system, i.e. the intersection of the class of diseases subsumed by *AcuteDigestiveSystemDisorder* with all those subsumed by *AcuteInflammatoryDisease*. Relation symbols begin with lower case, e.g. *hasAssociatedMorphology*. Roles are formed by a quantifier (here only the existential quantifier, $\exists$, is used), a relation symbol, followed by a dot and the name of a semantic category. For example, $\exists$ *hasAssociatedMorphology.Inflammation* denotes the semantic category whose instantiation is the set of all individuals related to an instance of *Inflammation* by the relation *hasAssociatedMorphology*. We can therefore rewrite the role group 1 and 2 entries in Table 2 (see Formula 1 in Fig. 1).

Looking at the parents, *RemovalOfForeignBodyFromStomach* and *IncisionOfStomach*[b], the first is an *is-a* descendent of *RemovalProcedure*, the latter is an *is-a* descendent of *IncisionProcedure*. Consequently, all instances of *RemovalOfForeignBodyFromTheStomachByIncision* are instances of both *IncisionProcedure* and

[b] as contained in the January 2006 SNOMED® CT sources

$$
\begin{aligned}
&\mathit{RemovalOfForeignBodyFromTheStomachByIncision\ IMPLIES}\\
&\quad \exists\, \mathit{rg}.(\exists\ \mathit{hasProcedureSite.StomachStructure\ AND}\\
&\qquad \exists\, \mathit{hasMethod.IncisionAction})\ \mathit{AND}\\
&\quad \exists\, \mathit{rg}.(\exists\ \mathit{hasProcedureSite.StomachStructure\ AND}\\
&\qquad \exists\, \mathit{hasDirectMorphology.ForeignBody\ AND}\\
&\qquad \exists\, \mathit{hasMethod.RemovalAction}) \qquad (1)
\end{aligned}
$$

Fig. 1 Formula 1

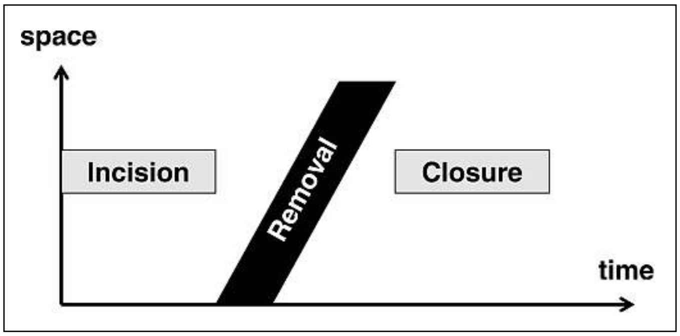

Fig. 2 Graphical representation of the process "Removal of Foreign Body from the Stomach by Incision"

RemovalProcedure, which must inherit the properties of both *Incision* and *Removal.* This is unreasonable because objects would be equally incised and removed. On the contrary, in a strict upper level ontology *Incision* and *Removal* are expected to be mutually exclusive. In reality, the surgeon first performs the incision and then the removal: *Incision* and *Removal* are two separate subprocedures and so are properly not **parents** but rather **parts**[c] of *RemovalOfForeignBodyFromTheStomachByIncsion.*

Figure 2 gives a graphic outline of this procedure which begins with the incision of the wall of the stomach, followed by the removal action and the closure of the wound (the latter is not mentioned in the procedure definition). These time-dependent sub-procedures stand to the main procedure in a *part-of* relationship. This is concordant with the commonly accepted mereological (*part-whole*) view of actions and processes, which are, according to Simons [9], characterized by time-dependent parts. Having this in mind it seems straightforward to re-interpret the relationship group attribute *rg* in (1) as the mereological primitive *hasPart* (see Formula 2 in Fig. 3).

Similarly, for the taxonomic parents we obtain (3) and (4) (see Fig. 3).

Looking up the SNOMED® CT hierarchy, we obtain exactly these definitions after replacing *rg* by *hasPart*. At a first glance this seems strange, since the main rationale for relationship groups, *viz.* the avoidance of ambiguities, makes no sense, here. Names such as *IncisionOfStomach*, suggest definitions without the *hasPart* role (see Formula 5 in Fig. 3).

The semantic difference is the following: Whereas *IncisingAStomach* denotes the atomic procedure of performing an incision onto a stomach, SNOMED® CT's *IncisionOfStomach* subsumes any complex procedure during which an incision of stomach is being performed. Analogously, *RemovalOfForeignBodyFromDigestiveSystem* subsumes any complex procedure during which a foreign body is extracted from the digestive system.

Looking still higher up the SNOMED® CT hierarchy, *IncisionOfStomach* is a child of *IncisionProcedure*, which is itself related to an *IncisionAction* by the relation *hasMethod* (see Formula 6 in Fig. 3).

As indicated above, the semantics of *rg* may be improved to derive (7) (see Formula 7 in Fig. 3).

An *IncisionProcedure* is, therefore, any procedure which has a part characterized by the enactment of an *Incision*. Only this broader definition justifies *IncisionProcedure* being the ancestor of nearly 1000 semantic categories: so many distinct flavors of incision do not exist, but more than 1000 surgical procedures have an incision as part of their description.

The example above is taken from the "procedure" branch of SNOMED® CT. Analogously, we give an example from the "disease/disorders" branch: *PepticUlcer-*

$$\begin{aligned}&\mathit{RemovalOfForeignBodyFromTheStomachByIncision\ IMPLIES}\\&\quad \exists \mathit{hasPart}.(\exists \mathit{hasProcedureSite.StomachStructure\ AND}\\&\qquad \exists \mathit{hasMethod.IncisionAction})\ \mathit{AND}\\&\quad \exists \mathit{hasPart}.(\exists \mathit{hasProcedureSite.StomachStructure\ AND}\\&\qquad \exists \mathit{hasDirectMorphology.ForeignBody\ AND}\\&\qquad \exists \mathit{hasMethod.\ RemovalAction})\end{aligned} \tag{2}$$

$$\begin{aligned}&\mathit{IncisionOfStomach\ IMPLIES}\\&\quad \exists \mathit{hasPart}.(\exists \mathit{hasProcedureSite.StomachStructure\ AND}\\&\qquad \exists \mathit{hasMethod.IncisionAction})\end{aligned} \tag{3}$$

$$\begin{aligned}&\mathit{RemovalOfForeignBodyFromStomach\ IMPLIES}\\&\quad \exists \mathit{hasPart}.(\exists \mathit{hasProcedureSite.StomachStructure\ AND}\\&\qquad \exists \mathit{hasDirectMorphology.ForeignBody\ AND}\\&\qquad \exists \mathit{hasMethod.RemovalAction})\end{aligned} \tag{4}$$

$$\begin{aligned}&\mathit{IncisingAStomach\ IMPLIES}\\&\quad \exists \mathit{hasProcedureSite.StomachStructure\ AND}\\&\quad \exists \mathit{hasMethod.IncisionAction}\end{aligned} \tag{5}$$

$$\mathit{IncisionProcedure\ IMPLIES}\ \exists \mathit{rg}.(\exists \mathit{hasMethod.IncisionAction}) \tag{6}$$

$$\mathit{IncisionProcedure\ IMPLIES}\ \exists \mathit{hasPart}.(\exists \mathit{hasMethod.IncisionAction}) \tag{7}$$

Fig. 3 Formulae 2-7

[c] Entity *A* has *B* as part is equivalent to the DL expression *A IMPLIES* $\exists$ *has-part.B.*

WithHaemorrhageAndObstruction which is subsumed by both *Gastrointestinal-Obstruction* and *PepticUlcerWith-Haemorrhage* (in the above sense). Table 3 shows the SNOMED® CT representation of this disease using relationship groups.

While the "*morphologic abnormalities*" *Obstruction* and *BleedingUlcer* represent pathologically altered body structures, the parents *GastrointestinalObstruction* and *PepticUlcerWithHemorrhage* represent more or less complex disorders characterized by having a part where the associated morphology is *Obstruction* or *Haemorrhage* respectively. A gastrointestinal obstruction may be complicated by perforation, gangrene, peritonitis, haemorrhage or not; in any of these cases, the obstruction is an integral part of the disease. This can be described using the *hasPart* relation (see Formula 8 in Fig. 4).

Similarly, a *PepticUlcerWithHaemorrhage* can be associated with an inflammatory process or with an injury – we may then formulate as can be seen in (9)(see Fig. 4).

Consequently, a *PepticUlcerWith-HaemorrhageAndObstruction* is a child of *GastrointestinalObstruction* as well as of *PepticUlcerWithHaemorrhage*. It is characterized by the existence of both *Obstruction* and *BleedingUlcer* as essential components of the whole entity (see Formula 10 in Fig. 4).

Another example: *TetralogyOfFallot*, a complex birth defect of the heart, is characterized by four cardinal manifestations (right ventricular hypertrophy, overriding of the aorta, pulmonary stenosis, and ventricular septal defect). Spackman et al. use this example to illustrate the need and benefit of relationship groups: grouping of related attribute-value pairs is possible while keeping the terminology model rather simple and still permitting description logic to make correct inferences about hierarchy placement and subsumption [5] (see Formula 11 in Fig. 4).

Replacing *rg* with *hasPart* leads to a semantically clearer representation (here: the four cardinal manifestations are components – parts – of the complex pathological state) while the advantages already mentioned are retained. *TetralogyOfFallot-hasParts* (is characterized by the existence of) *PulmonaryStenosis* AND *Hypertrophy-OfRightVentricle* AND *Incomplete-*

Table 3 Entries in the January 2006 SNOMED® CT (UK release) core relationships table for "*PepticUlcerWithHaemorrhageAndObstruction*"

SNOMED® CT Concept 1	SNOMED® CT Relationship	SNOMED® CT Concept 2	RG
PepticUlcer With Haemorrhage AndObstruction	*Is A*	*GastrointestinalObstruction (disorder)*	*0*
	Is A	*PepticUlcerWithHaemorrhage (disorder)*	*0*
	causativeAgent	*GastricAcid (substance)*	*0*
	associatedMorphology	*Obstruction (morphologic abnormality)*	*1*
	Finding site	*GastrointestinalTractStructure (body structure)*	*1*
	associatedMorphology	*BleedingUlcer (morphologic abnormality)*	*2*
	Finding site	*UpperGastrointestinalTract Structure (body structure)*	*2*

$$\begin{aligned}&\mathit{GastrointestinalObstruction}\ \ \mathit{IMPLIES}\\&\quad \exists \mathit{hasPart}.(\exists \mathit{hasFindingSite}.\mathit{GastrointestinalTractStructure}\ \ \mathit{AND}\\&\quad \exists \mathit{hasAssociatedMorphology}.\mathit{Obstruction})\end{aligned} \tag{8}$$

$$\begin{aligned}&\mathit{PepticUlcerWithHaemorrhage}\ \ \mathit{IMPLIES}\\&\quad \exists \mathit{hasPart}.(\exists \mathit{hasFindingSite}.\mathit{UpperGastrointestinalTractStructure}\ \ \mathit{AND}\\&\quad \exists \mathit{hasAssociatedMorphology}.\mathit{BleedingUlcer})\end{aligned} \tag{9}$$

$$\begin{aligned}&\mathit{PepticUlcerWithHaemorrhageAndObstruction}\ \ \mathit{IMPLIES}\\&\quad \exists \mathit{hasPart}.(\exists\ \mathit{hasFindingSite}.\mathit{GastrointestinalTractStructure}\ \ \mathit{AND}\\&\qquad\quad \exists \mathit{hasAssociatedMorphology}.\mathit{Obstruction})\ \ \mathit{AND}\\&\quad \exists \mathit{hasPart}.(\exists\ \mathit{hasFindingSite}.\mathit{UpperGastrointestinalTractStructure}\ \ \mathit{AND}\\&\qquad\quad \exists \mathit{hasAssociatedMorphology}.\mathit{BleedingUlcer})\end{aligned} \tag{10}$$

$$\begin{aligned}&\mathit{TetralogyOfFallot}\ \mathit{IMPLIES}\\&\quad \exists \mathit{rg}.(\exists\ \mathit{hasFindingSite}.\mathit{PulmonaryValve}\ \mathit{AND}\\&\qquad \exists \mathit{hasAssociatedMorphology}.\mathit{Stenosis})\ \mathit{AND}\\&\quad \exists \mathit{rg}.(\exists\ \mathit{hasFindingSite}.\mathit{RightVentricle}\ \mathit{AND}\\&\qquad \exists \mathit{hasAssociatedMorphology}.\mathit{Hypertrophy})\ \mathit{AND}\\&\quad \exists \mathit{rg}.(\exists\ \mathit{hasFindingSite}.\mathit{InterventricularSeptum}\ \mathit{AND}\\&\qquad \exists \mathit{hasAssociatedMorphology}.\mathit{IncompleteClosure})\ \mathit{AND}\\&\quad \exists \mathit{rg}.(\exists\ \mathit{hasFindingSite}.\mathit{Aorta}\ \mathit{AND}\\&\qquad \exists \mathit{hasAssociatedMorphology}.\mathit{Overriding})\end{aligned} \tag{11}$$

Fig. 4 Formulae 8-11

Table 4 Entries in the January 2006 SNOMED® CT core relationships table for "*UrgentScrapeOfTheLeftCornea*"

SNOMED® CT Concept 1	SNOMED® CT Relationship	SNOMED® CT Concept 2
UrgentScrapeOfTheLeftCornea	*Is A*	*TissueSpecimenFromEye*
	Is A	*SpecimenFromCornea*
	Is A	*Scrapings*
	specimenSourceTopography	*CornealStructure*
	specimenSourceTopography	*BodyTissueStructure*
	specimenProcedure	*Scraping*
	priority	*Urgent*
	laterality	*Left*

$$
\begin{aligned}
&\mathit{UrgentScrapeOfLeftCornea\ IMPLIES}\\
&\quad \mathit{Specimen\ from\ Cornea\ AND}\\
&\quad \exists rg.(\exists specimenSourceTopography.CornealStructure\ AND\\
&\qquad \exists Laterality.Left)\ AND\\
&\quad \exists rg.(\exists specimenProcedure.Scraping\ AND\ \exists Priority.Urgent) \qquad (12)
\end{aligned}
$$

Fig. 5 Formula 12

ClosureOfInterventricularSeptum AND *OverridingAorta*.

However, there may be situations in SNOMED® CT where the interpretation of *rg* as *hasPart* would not be correct. For example, within the current SNOMED® CT content, it is possible to construct the post-coordinated composition of an *UrgentScrapeOfTheLeftCornea*, by adding the *Urgent* and *Left* qualifiers as appropriate. The resulting flattened (not role grouped) representation is shown in Table 4.

Here it would be impossible to tell whether the attribute *Left* should be applied to the swab or the cornea. So, we might want to formalize (see Formula 12 in Fig. 5).

4. Conclusion

An analysis of relationship groups in SNOMED® CT revealed counter-intuitive inference results. Our suggestions seem to improve SNOMED® CT in terms of increased adequacy and avoidance of inconsistencies, while they require only minor modifications of SNOMED® CT's architecture:

- In the SNOMED distribution, rename the relationship group attribute *rg* by *hasPart* or *hasSubprocess* where it appears between a complex process and its subprocesses (i.e. especially in the disease and procedure chapters of SNOMED® CT).
- Make a clearer distinction between atomic categories (such as *IncisionAction or Inflammation/MorphologicAbnormality*) and those entities which have atomic categories as parts (such as *IncisionProcess or InflammatoryDisorder*). The present names are misleading, especially for procedures.

Finally, one has to take into account that there are scenarios in which the use of relationship groups seems adequate, without, however, corresponding to a mereological relation. A more detailed ontological inquiry of these cases is still due.

Acknowledgments
This work was supported by the *EU Network of Excellence Semantic Interoperability and Data Mining in Biomedicine* (NoE 507505), cf. http://www.semanticmining.org. We also thank Ulrike Sattler (Manchester, UK) and Kent Spackman (Portland, OR, USA), for their helpful discussions.

References

1. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 1998: 37(4/5): 394-403.
2. Spackman KA, Campbell K, Cote RA. SNOMED RT: A reference terminology for health care. In: Masys DR, editor. AMIA'97 – Proceedings of the 1997 AMIA Annual Fall Symposium. Philadelphia, PA: Hanley & Belfus; 1997. pp 640-4.
3. Smith B. Beyond concepts, or: Ontology as reality representation. In: Varzi AC, Vieu L, editors. FOIS 2004 – Proceedings of the 3rd International Conference on Formal Ontology in Information Systems. Amsterdam: IOS Press; 2004. pp 73-84.
4. Spackman KA, Reynoso G. Examining SNOMED from the perspective of formal ontological principles: Some preliminary analysis and observations. In: Hahn U, Schulz S, Cornet S, editors. KR-MED 2004 – Proceedings of the 1st International Workshop on Formal Biomedical Knowledge Representation, Whistler, B.C., Canada, June 1, 2004. Bethesda, MD: American Medical Informatics Association (AMIA), 2004. pp 72–80. Published via http://CEUR-WS.org/Vol-102/.
5. Baader F, Calvanese D, McGuinness D, Nardi D, Patel-Schneider P, editors. The Description Logic Handbook. Theory, Implementation and Applications. Cambridge, UK: Cambridge University Press; 2003.
6. Spackman KA, Dionne R, Mays E, Weis J. Role grouping as an extension to the description logic of ONTYLOG, motivated by entity modeling in SNOMED. In: Kohane IS, editor. AMIA 2002 – Proceedings of the Annual Symposium of the American Medical Informatics Association, Philadelphia, PA: Hanley & Belfus, 2002. pp. 712-6.
7. Spackman KA, Campbell KE. Compositional concept representation using SNOMED: Towards further convergence of clinical terminologies. In Chute CG, editor, AMIA '98 – Proceedings of the 1998 AMIA Annual Fall Symposium, Philadelphia, PA: Hanley & Belfus; 1998. pp 740-4.
8. Unified Medical Language System (UMLS). Bethesda, MD: National Library of Medicine, 2006.
9. Simons P. Parts: A Study in Ontology. Oxford: Clarendon Press; 1987.

Correspondence to:
Stefan Schulz
Abteilung Medizinische Informatik
Universtätsklinikum Freiburg
Stefan-Meier-Str. 26
79106 Freiburg
Germany
E-mail: stschulz@uni-freiburg.de
http://www.imbi.uni-freiburg.de/medinf/~schulz.htm