2005-10-06

Concept Systems and Ontologies

Recommendations based on discussions between realist philosophers and ISO/CEN experts concerning the standards addressing "concepts" and related terms

Gunnar O. Klein, *CEN / TC 251 chairman*

Centre for Medical Terminology, Karolinska Institutet, SE- 171 77 Stockholm, Sweden
gunnar.klein@ki.se

Barry Smith*, professor*
Department of Philosophy, University at Buffalo, Buffalo, NY 14260, USA
Institute for Formal Ontology and Medical Information Science, Saarland University, 66041 Saarbrücken, Germany
phismith@buffalo.edu

1. Foreword

This is the third draft of a paper that aims to clarify the apparent contradictions in the views presented in certain standards and other specifications of health informatics systems, contradictions which come to light when the latter are evaluated from the perspective of realist philosophy.

One of the origins of this document was Klein's discussion paper of 2005-07-02 entitled "Conceptology vs Reality" and the responses from Smith, as well as the several hours of discussions during the 2005 MIE meeting in Geneva.

The structure of this paper is:

- An introduction which summarizes some of the background discussion papers and the issue of apparent contradictions
- A presentation of a definition of "concept" in one specific and consistent meaning that is adapted to the needs of terminological systems
- A presentation of proposed terms, with their definitions, designed to substitute for what we believe are misleading uses of the term "concept" in current literature.
- A discussion of relations between concepts
- A summary of terminology for ontologies as representations of the real world
- An introduction to the topic of representation of relations in ontologies
- A (partial) mapping between the world of concepts and the world of reality
- A presentation of some aspects of the proposal as it relates to the associations between concept systems, ontologies and databases including data on individual instances

This paper was financed by the European Union (FP6) Network of Excellence Semantic Interoperability and Data Mining in Biomedicine.

2. Introduction

During the last years the realist philosophers Barry Smith and Ingvar Johansson have been challenging the standardization work in Health Informatics as it deals with what are standardly called “concepts”. One recent reason to bring this up is the discussion on the first working draft of the CEN/TS: *Health informatics – Categorial structure for anatomy*, see also comments by Klein on this draft. Werner Ceusters, Anand Kumar and Cornelius Rosse have also contributed to these discusssions.

Barry Smith has written a series of papers in which he and co-workers argue that the concept of “concept” is used in much of the work on biomedical terminologies, ontologies and controlled vocabularies in ways that are seriously flawed.

See for instance: (1) Barry Smith, “Beyond Concepts: Ontology as Reality Representation” (FOIS 2004); http://ontology.buffalo.edu/bfo/Beyond_Concepts.pdf

Another paper of interest for those involved in terminology standardization as it applies to ISO/TC 37 (founded by Eugen Wüster) and to its successor institutions is the commentary (2) by Smith, Ceusters and Temmerman called ‘Wüsteria”, presented at MIE 2005; http://ontology.buffalo.edu/medo/Wuesteria.pdf

See also:
Barry Smith and Werner Ceusters, “Ontology as the Core Discipline of Biomedical Informatics: Legacies of the Past and Recommendations for the Future Direction of Research”, forthcoming in *Computing, Philosophy, and Cognitive Science*, Cambridge: Cambridge Scholars Press, 2006;
http://ontology.buffalo.edu/medo/Recommendations_2005.pdf

Barry Smith, “From Concepts to Clinical Reality: From Concepts to Clinical Reality: An Essay on the Benchmarking of Biomedical Terminologies”, *Journal of Biomedical Informatics*, forthcoming: http://ontology.buffalo.edu/medo/reasoningBT.pdf

We agree that the term “concept” has been misused in some writings, but hold that it can still serve an important role in modern health informatics. We agree also that there is a need to have a serious and constructive discussion on how to resolve the problems which have arisen through its misuse. This is an attempt to propose a set of alternative terms and definitions that may replace the single term “concept” in certain specialist contexts, where information structures and semantic interoperability is the concern. We have of course no ambition to change the very frequent and varying uses of the term “concept” in natural language discourse.

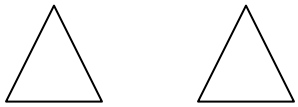
While one of us, Klein, is an M.D. and has his main function in Health Informatics – and hence the use of many examples from the health area – there is nothing specific to the healthcare domain about the problems and solutions discussed.

3. Concept – A strange animal with many heads

3.1 The history of the use of the term "concept"

The term "concept" and its cognates (*conceptus*, idea, notion, *Begriff*) has a long history, going back at least as far as Plato. For present purposes it is important to refer to the great medieval dispute between realists, conceptualists and nominalists over the so-called "problem of universals".

The **realists** hold that there are universals – invariant patterns (also called characters or essences) – existing on the side of entities in reality, and that it is in virtue of such universals that particulars – for example these two particular triangular shapes – manifest relations of similarity to each other:



According to the realist view such relations of similarity would exist even if there were no cognitive subjects in a position to observe them. Each single universal can be exemplified multiple times by an open-ended plurality of particulars. Universals are further organized into trees of universals of greater and lesser generality, called *genera* and *species*, respectively. All such universals are however distinguished from the particulars in reality which we also call "instances" (see further below on terminology for ontologies). When the realist philosophers use expressions like "general concepts" and "general terms" they relate to universals on the side of reality. The relation of similarity between a given set of instances exists because the *same universal* is exemplifed by each of the given shapes.

Conceptualists, in contrast, hold that there are no universals on the side of entities in reality, but rather only in our minds. One and the same general concept, say *triangle*, can be related to a plurality of triangles in reality. Different cognitive subjects can share the same general concepts, which are in this sense multiply exemplified in different minds. But single concepts do not correspond to single universals or invariants on the side of reality. Rather, all concepts relate to their instances in an *ad hoc* way, as in the case of concepts such as *things you might take on a holiday* or *things you might need to build a weapon.*

Nominalists, finally, hold that there are universals neither in reality nor in our minds, but rather only general terms. Nominalists thus deny the existence of general concepts which can be shared by a plurality of cognitive subjects. General terms are mere labels for *ad hoc* collections of things or events. When different subjects apply the same general term, say "triangle", to each of the two particulars depicted above, then, according to the nominalist their respective ideas have just as little objective similarity to each other as do the entities in reality to which these terms are applied.

3.2 Finding a solution by the use of separate terms

Traces of all three of these positions are present in contemporary uses of the term "concept" in terminology circles. Thus in some contexts the term refers to what would more properly be called "universals" in the sense of the realist doctrine; in some contexts it refers to general ideas in people's minds; and in yet other contexts the term refers merely to general terms in some controlled language. Contradictions then arise because these three distinct readings are not clearly distinguished in the terminologist literature.

Moreover, psychological, linguistic and computational uses of the term "concept" have in more recent years also been added to this mix, so that there are today a number of different valid viewpoints developed for different purposes where the term "concept" has been applied with very different meanings. The problems arise only where the term is used with a meaning that is left unspecified or where the term is used with different and contradictory meanings in one and the same text.

Instead of abandoning the term completely, as has been suggested in some circles, we offer here a careful analysis of the different meanings of "concept" and then propose separate terms and definitions to ensure disambiguation. At the same time we offer a series of warnings concerning standard ways of misusing the term "concept" and the problems to which they give rise.

We realize that the proposals may not be intuitively clear and will at first seem unfamiliar. At the same time however we are convinced that urgent reforms are required if the present problems are to be remedied in a timely manner. Those of our proposed replacement terms which contain "concept" as constituent may, in certain circumstances, be substituted for by "concept". Then, however, the author in question should be careful to provide in every case a clear indication that he is using "concept" as such a substitute. More importantly, when different meanings of "concept" are appealed to in the same context, then different terms must be used to convey the corresponding different meanings.

The main focus of this communication is systematic terminological work in fields like biomedicine, where the need arises to develop systematic representations of real world entities such as disorders or anatomical structures. Hence part of our goal is to support also the development of such systematic representations – ontologies – in such a way that they can be used for information integration and alignment and also for automatic reasoning.

Our proposal, simply put, is that "concept" refers always to the meaning of the corresponding term, as agreed upon by responsible persons. This means that even in those areas of health care informatics standardization where regulatory bodies to unify terminology as it relates to clearly human social constructions such as the "mandate" in ENV 13940 the use of the term "concept" will need to be subject to reform. This is because even where we are discussing human constructions such as mandates, agreements, contracts and the like, there is a distinction between the entity referred to on the one hand (the mandate in question) and the meaning of the term ("mandate") which we are using to refer to this entity.

4 Recommendations

We recommend that the following terms should be used exclusively, in terminological contexts, in the following ways.

4.1

concept

concept in a terminological system	(synonym 1)
agreed meaning of a term	(synonym 2)
meaning of a term as agreed upon by a group of responsible persons	(definition)

Note 1: This is related to the definition in the current ISO 1087-1 from the year 2000: “unit of knowledge created by a unique combination of characteristics”. However, we prefer “meaning” to “unit of knowledge”, since the latter has several problems. Thus there can be agreed meanings for terms like “unicorn”, but such meanings do not correspond to any unit of knowledge, since there are no corresponding entities in reality about which knowledge could be gained. We also prefer “meaning” to “unit of thought”, in order to avoid the psychological connotations of the latter. We postpone to a later communication discussion of the term “characteristic”, which is subject to the same sorts of problems as have affected the term “concept”.

Note 2: The fact that an identified group of responsible persons (the persons responsible for a given terminological system) share a common understanding of the meaning of a term – typically captured by means of a definition – is important for many of the contexts where the term “concept” is used today in modern informatics. This is what distinguishes concepts, as we shall here understand them, from individual ideas in the minds of cognitive subjects. In particular, the existence of agreed meanings on the part of responsible persons is clearly important for the development of formalized standards, including international standards, but it is important also for other groups responsible for a single terminology.

Note 3: Philosophers differ as to what is meant by “meaning”, sometimes it is expressed as: that which remains constant when a word in one language is translated correctly into another language. More operational interpretations of “meaning” in this context include:

Given cognitive subjects demonstrate that they have command of the same meaning of a term when they share the ability, upon receiving information containing this term, to associate it with exactly the same referents.

Meaning in this sense is closely related to, but importantly distinct from, definition. Thus the term “concept”, on the reading “agreed meaning of a term”, does not refer to a corresponding definition, or to any other specific representation of this meaning in some natural or artificial language or in some formal model. One and the same concept (in the sense of “agreed meaning”) will typically correspond to several alternative ways of expressing this meaning, and thus to different linguistic expressions in the same as well as in different languages.

Note 4: Meanings must in this context be specified, and there are two possible ways to do

this in a terminological system: by the linguistic representation of a definition or elucidation, or by a translation of the corresponding general term into a different language, where in the specific context of the terminological system in question the provision of alternative terms is able to convey a sufficient understanding of the meaning in question. Usually the representation of a definition in the form of a statement of necessary and sufficient conditions is preferable. However, sometimes we can only use *ostensive* definitions. E.g. SARS = that syndrome which a certain defined set of individual patient cases have in common. (Note the way in which an ostensive definition of this sort points to a universal in reality, i.e. to a certain multiply exemplified entity.)

4.2

concept definition

concept definition representation (synonym1)

definition (synonym2)

representation of a concept (as agreed meaning of a term) by a descriptive statement or a formal expression which serves to differentiate it from related concepts (definition)

Note: This may also be referred to as just “definition”, but we believe that it may be important to use the longer term in order to clarify that it is a particular representation or formulation of a definition that is meant and not the abstract meaning as such. This may be important where there is more than one definition which captures the same meaning. Note that formal expressions, including graphical models such as UML or Venn diagrams, may also serve the definition of concepts by showing their interrelations with other concepts.

4.3
concept system

structured collection of representations of concepts including associated terms and relations (proposed definition)

Note: The concept system is a collection of elements (see 4.4) which are related together via interconnections representing relations between the corresponding meanings.

Note: While the definition of a concept (in the sense of: agreed meaning of a term) will most often take the form of a linguistic expression, such a definition may also be expressed by other means, in particular in ways involving appeal to relations between concept system nodes.

4.4
concept system node

information element within the structure of a concept system that is a pointer or place-holder linking one or several synonymous terms with a given **concept definition representation** and also used in the representation of relations between concepts

Note 1: In many modern concept systems designed for use with Information Technology, the concept system node has as a key component a numeric concept identifier. Where a concept system exists in graph-theoretical form, this can be thought of as a node or vertex of the graph. The edges of the graph then represent relations between the concepts represented by the corresponding nodes.

Note 2: When SNOMED CT, for example, uses the term "concept", then what it means is "concept system node" in the terminology advanced here. Certainly SNOMED officially defines "concept" as meaning: "unit of thought". When we inspect its actual practice, however, then we discover that concepts in SNOMED CT are used as pointers which allow the capturing of relationships of synonymy between terms (there called "descriptions") and also of associations with other concept nodes as well as with related attributes. This is also the most common meaning of "concept" in information model standards for Electronic Health Records or other health informatics standards such as the HL7 RIM or CEN message standards.

4.5 Psychological uses of the term "concept"

In natural language, and in some of Wüster's work, the term "concept" is used to mean what would more properly be called "mental concept" or "idea" ("noesis" in Aristotelian terminology), defined as: a certain state of the brain, for example as associated with the use of a general term.

This is not an important connotation for purposes of standardization in the domain of terminology systems. We mention it only in order to point out the need to separate this very common use of "concept" in natural language and many scientific texts from the meaning recommended in the context of terminology. This need is all the more urgent given the influence of Wüster's ideas on the literature of terminology standardization.

Note that our concerns here do not concern areas like conceptual psychology, where the term "concept" can of course continue to be used but in a different meaning.

5 Relations between concepts

ISO 1087-1 and many other works on concepts specify a number of different types of relations between concepts. If a concept is the meaning of a term and not a real world entity which may correspond to or be referred to by the concept, then there is really only one relation which should be used, namely the generic *is_a* relation which exists between superordinate and subordinate concepts. In ISO 1087-1 this is defined as:

generic relation

relation between two **concepts** (3.2.1) where the **intension** (3.2.9) of one of the **concepts** (3.2.1) includes that of the other **concept** (3.2.1) and at least one additional **delimiting characteristic** (3.2.7)

This generic relation between concepts can be called a **semantic relation** to stress the fact that it is a relation which has agreed meanings as its relata. In what follows we shall refer to it by means of the phrase "*is_a* (is narrower in meaning than)".

What ISO1087-1 calls "partitive" and "associative relations" (such as *part_of* or *causes*) are not appropriate for concept systems – since they are not relations which hold between meanings. Such relations should be used, rather, in ontologies (see below), in which real world entities are taken into account.

6 Terminology for ontologies

6.1 Entities

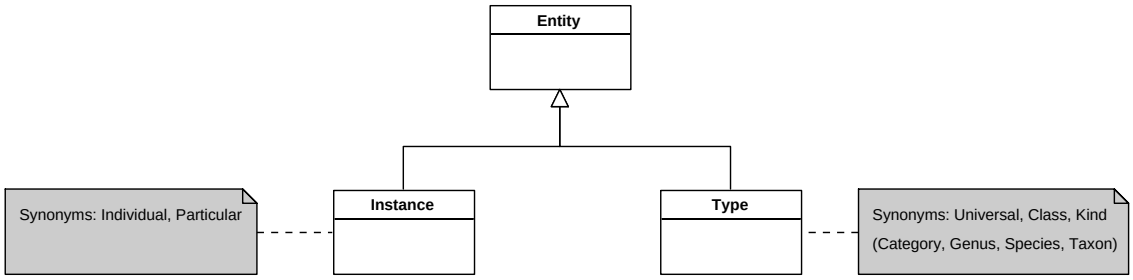
If concept systems are systems of meanings, we need a supplementary terminology for those representation systems which relate to real world entities, both those investigated by the natural sciences (such as cells or tumours) and those existing in administrative domains (such as mandates or documents recording lab results). Unfortunately this terminology – the terminology of ontology – is not yet established in a consistent way in informatics and terminology circles.

What we propose here reflects an emerging consensus in ontological research; but we also provide alternative synonyms (in parentheses) to serve as guidance for the wider community.

The entities in reality are of two kinds: "instances" and "types".

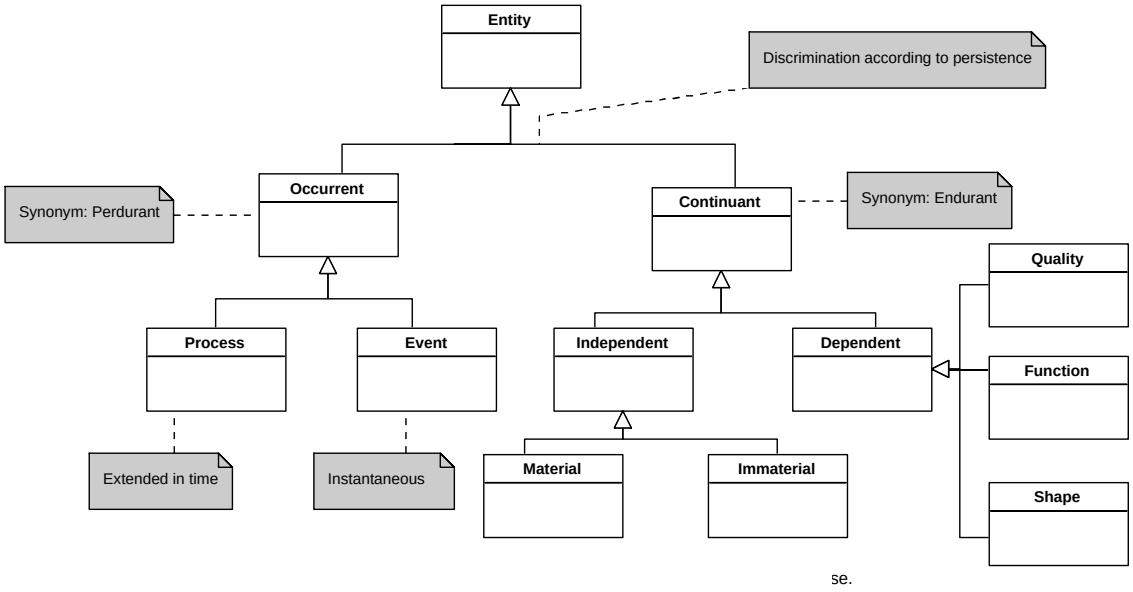
What we here propose to call "instances" have in the ontology literature also been called "individuals" or "particulars". These terms can be regarded as synonyms in this context, but we prefer the term "instance" since it draws attention to the fact that the entities in question are *instances of* corresponding *types*. Thus the particular cell in this Petri-dish is an *instance of* the type *cell*, of the type *B-lymphocyte*, and so on.

Alternative terms for what we here call "type" (which itself corresponds to what the medieval realists called "universal") are: "class", "kind" (also sometimes "category", "*genus*", "*species*", "*taxon*").



Entitities (instances and types) can be further classified according to the following scheme focusing on their persistence, dividing them in two main kinds”

Occurrent (perdurant)	**Continuant** (endurant)				
(process, event)	independent		dependent		
	material	immaterial	quality	shape	
Examples actions	cells	body cavities	temperature	triangularity	



6.2 Relations between real world entities

Individual instances can have various relations to other instances. For example Mary's heart is part of Mary; Mary's run is part of Mary's morning work-out, and so on. In some cases *all* instances of a given type stand in such relations to correlated instances of some other type. Thus all instances of the type *human* are also instances of the type *mammal.* All instances of the type *adult* are identical to some instance of the type *child* existing at an earlier time. All instances of the type *nucleus* are adjacent to some instance of the type *cytoplasm*, and so on. Such relations are universal, in the sense that they hold of all instances of a given type (namely, in each case, of the type first mentioned). They often do not hold when inverted (thus it is not the case that all instances of the type *cytoplasm* are adjacent to some instance of the type *nucleus*).

It is such universal relations between types which are properly captured in an ontology, which is a representation of the types exemplified by instances in a given domain of reality, and of the universal relations obtaining between such types.

For supporting arguments see:
Barry Smith, Werner Ceusters, Bert Klagges, Jacob Köhler, Anand Kumar, Jane Lomax, Chris Mungall, Fabian Neuhaus, Alan Rector, Cornelius Rosse, "Relations in Biomedical Ontologies", *Genome Biology*, 2005, 6 (5), R46; http://genomebiology.com/2005/6/5/R46

Much of biomedical knowledge, for example as contained within textbooks, is about the systematisation of the universal relations between types of real world entities. Data in medical records, on the other hand, is often a matter of instance-level relations between the corresponding instances, for instance data to the effect that traces of this chemical are

located in this blood sample here and now. Where designations of types (for example via clinical codes) are used in expressions of such data, then this is to classify the corresponding instances.

7 Correspondence between concept systems and ontologies

Some concepts – and very many of those concepts used in biomedical terminology systems (more properly called "concept systems" in light of the above) – have a relation to entities in the real world which is closely analogous to the relation between types and their instances. Mary is an *instance of* the type *human being*, but Mary also *falls under* the concept *human being.* The type *human being* stands in an *is_a* (is a subtype of) relation to the type *mammal.* But the concept *human being* stands in an *is_a* (is narrower in meaning than) relation to the concept *mammal.*

In spite of this parallelism, however, the term "concept" should still never be used in place of "type" ("universal", "kind") as thus defined, because the parallelism is only partial. First, there are concepts such as

unicorn
case of pneumonia in Russian fiction
fractured lip

which correspond to no real world entities on the instance level. Second, there are concepts such as

non-rainy day
non-mammal
relative of possible smoker
other metalworker in New Zealand
person admitted before 9a.m.
mixture of water and alcohol containing zero amount of alcohol

which correspond to no real-world entities on the level of types.

Thus only some subset of the nodes in a given biomedical concept system will be mappable in a 1-1 way to corresponding nodes in a type system (ontology).

Type and *concept* should be kept clearly separate also because of the different ways in which they are connected by relations.

Certainly some of the nodes in concept systems stand in *is_a* relations (is narrower in meaning than) in ways which are isomorphic to the *is_a* relations (is subtype of) which hold between the corresponding nodes in the ontology. But there are many more *is_a* relations of the former type than of the latter. There are also many non-*is_a* (is subtype

of) relations connecting types in ontologies which have no application to concept systems, including all the familiar relations *part_of*, *transformation_of* , *located_in*, *derives_from*, *adjacent_to*, *participates_in*, and so on.

Concept systems are thus simple hierarchies, whose nodes are joined together exclusively by *is_a* (is narrower in meaning than) relations. Ontologies typically manifest much more complex graph-theoretic structures, in which many further relational edges are included. On the other hand concept systems may be much richer, since they may include many nodes which correspond to no universals on the side of reality.

The following figure illustrates the different worlds of concept systems, ontologies and the information models.