The Importance of Being Earnest about Definitions

Susan Thomas

SAP Research
SAP AG
Vincenz-Priessnitz-Str. 1
76131 Karlsruhe, Germany
susan.marie.thomas@sap.com

Abstract: Ideas from terminology management, the science of terms and definitions, can be used to improve the quality of software and data models, as well as to facilitate the achievement of some of the current goals of computer science like the realization of the Semantic Web and the re-use of software components. This paper explains the principles and methods of terminology management and relates it to problems in computer science. The predominant aim of the paper is to raise awareness of both the importance of definitions and the difficulties associated with them in order to pave the way for the effective use of terminology management in software projects.

1 Motivation

Better skills in the art of writing definitions of terms (words or phrases) can greatly facilitate the achievement of some of the current goals of computer science like the realization of the Semantic Web and the standardization and re-use of components of all kinds, including data models. But, among engineers there is a general neglect of this art, which is seen as "just words" or "just documentation". And we engineers are widely and not implausibly believed to have said on receiving our degrees: *I always wanted to be an engineer and now I are one*. However, since definitions are more about meaning than grammar, there is still some hope that the ideas in this paper will find an audience.

The rest of the paper is organized as follows. First, there is a section about the importance and role of definitions in some topical Information Technology (IT) projects. Then some essential concepts related to definitions are presented. This is followed by a section on how techniques from terminology management — the discipline concerned with terms and definitions — are transferable to IT, followed by a closely related section which gives examples of definitions. Finally, there is a short outlook section about future work.

The predominant intention and contribution of the present paper is to raise awareness of the importance and difficulties associated with meaning and definitions, and to relate these difficulties to some common problems in IT. Such awareness could help improve the quality of IT products.

2 The importance of definitions in information technology

Three well-known projects in which definitions play a prominent role are: WordNet, the Semantic Web and ebXML Core Components, which is related to inter-operability and data integration.

2.1 WordNet lexical database

WordNet is an electronic lexical database of English, which was created by hand. Initially, it contained no definitions, but as discussed later, this decision turned out to be a big mistake and was reversed. WordNet has been used in numerous projects in Natural Language Processing and Artificial Intelligence. Having been under development since the ‘80s, it now has a coverage comparable to the popular Meriam-Webster® dictionary. But, it is organized along entirely different lines. Rather than being arranged alphabetically by word, it is a network of lexical and semantic relationships, which means relationships directly between words, and relationships between the meanings of words, respectively.

Semantics like many of the words and concepts used in this paper is derived from ancient Greek language and philosophy. It is a 5-dollar word for ‘meaning’, and has become rather over-used recently in IT. Its usual home is in linguistics, where it is part of various theories of language that distinguish between syntax, semantics and pragmatics among other things. In fact, there is a long-standing philosophical dispute about the meaning of *meaning*, and there are various theories of meaning, but that is a subject for another paper. As with all philosophical questions, the aim of discussion is not to come to a correct answer, but to become aware of and to articulate one’s own intellectual and value commitments to see where they lead, and to see if this is where you want to go.

Meanings or concepts in WordNet are represented by sets of words with essentially the same meaning, that is, *synonyms*. These sets are called *synsets*. This is another interesting contrast with standard dictionaries, in which the point is to distinguish between the often subtle shades of meaning of synonyms as shown by this example from the on-line American Heritage® dictionary[1]:

---

[1] Web site: http://www.bartleby.com/

"SYNONYMS: *perceptible, palpable, appreciable, noticeable, discernible* These adjectives apply to what is capable of being apprehended as being real by the mind or through the senses. *Perceptible* is the least specific: *a perceptible pause in the flow of his speech*. *Palpable* applies both to what is perceptible by means of the sense of touch and to what is readily perceived by the mind: "*The advantages Mr. Falkland possessed are palpable*" (William Godwin). What is *appreciable* is capable of being estimated or measured: *dumping appreciable amounts of noxious waste into the harbor*. *Noticeable* means easily observed: *noticeable shadows under your eyes*. *Discernible* means distinguishable, especially by the faculty of vision or the intellect: *no discernible progress in the contract negotiations*."

It is doubtful that any synset fully passes the synonym test attributed to Leibnitz: a word is a synonym of another if replacing it in a sentence does not change the meaning of the sentence. However, a very good example of real synonymous use of three words in one sentence comes from a passage in the film *Annie Hall* by Woody Alan. Woody plays a Manhattan neurotic who goes to California and casually remarks that of course it's clean in Hollywood compared to New York because in Hollywood they turn all of their garbage into television shows. Shortly after this, he's at a pretentious show-biz party where he overhears a strangely-dressed man saying:

> *"Right now it's only a* ***notion****, but I think I can get money to make it into a* ***concept*** *... and later turn it into an* ***idea****."*

Deciding when to differentiate between words, and when to group them together into a synset is far from trivial, and there is no correct answer. For example, WordNet does not put *notion, concept* and *idea* into one synset. Nonetheless, WordNet seems to have resolved these problems well enough to have been useful in a number of IT projects.

In WordNet the different categories of words: nouns, verbs, adjectives and adverbs are organized in different ways. Noun synsets are organized into a *taxonomy*, which essentially means a *class hierarchy,* which means *holy rule*. An extract of a hierarchy is shown next. Curly brackets enclose synsets. The symbol @-> denotes is-a-kind-of.

> {robin, redbreast}@->{bird}@->{animal, animate_being}@->{organism, life_form, living_thing} (example from [Fe98] p. 25)

In addition to being organized taxonomically, nouns can be structured into part/whole hierarchies if applicable. A less fundamental, but interesting relationship between nouns, is that of *antonymy,* meaning *opposition*. *Victory* and *defeat* are antonyms.

Verb synsets are also organized into hierarchies based on a few relationships like *troponymy*, meaning *way or manner*. *To march* is a troponym of *to walk* because marching is a way or manner of walking.

Adjectives have a rich organizational structure based on antonymy and similarity, and adverbs are linked to adjectives from which they are derived.

Returning to the subject of definitions, initially WordNet contained no definitions of concepts/synsets at all. The creators of WordNet thought that the meanings of words could be easily inferred from the semantic and lexical relations captured in the database. Definitions were, therefore, considered to be superfluous. However, as the database grew they soon discovered the extreme difficulty of understanding meanings without the help of definitions and illustrative uses of words. So, they started adding them. As they themselves admit in [Fe98], page Roman numeral xx, they learned the hard way the importance of definitions and acquired greater respect for traditional lexicographers.

2.2 The Semantic Web

The authors of [Fr04] in writing about the relationship between the W3C's Semantic Web project and the OMG's Model Driven Architecture ® (MDA) initiative say that "the Semantic Web, as envisioned by Tim Berners-Lee and many others since, is a logical extension of the current Web that enables explicit representations of term meanings". They go on to say that explicit descriptions of terms and their interrelationships enable programs to "understand" the meanings of terms. These machine-processable definitions are encoded in *ontologies,* which means the *science of being*. Ontologies are written in formal languages, usually based on a subset of first-order logic.

However, it is important to note that ontologies supplement, but do not replace human-readable versions of information. This view is confirmed by the experience of the creators of WordNet. As they learned the hard way, terms and their interrelationships encoded for machine processing, ontologies in the context of the Semantic Web, are not a substitute for standard definitions and illustrative examples. Developers of the Semantic Web can directly profit from the experience of the WordNet creators. Some ontology developers like the authors of The Enterprise Ontology [Us98] do, in fact, advocate starting the development of an ontology with natural-language definitions of key terms. There are at least two compelling reasons to have such definitions: first, to enable users to understand the system; and secondly, to enable developers to understand it in order to make changes. Terminology management, which is explained later, provides a methodical way to start the development of ontologies, and to write definitions.

2.3 ebXML Core Components, business inter-operability and data integration

The idea of Core Components was developed as part of the ebXML Framework, and is a new approach to solve the problem of data interoperability in e-business. It specifies a method to develop re-usable building blocks that represent the data commonly used in business today. Such building blocks, or components, range in size from generic data-types to whole business documents like purchase orders.

The goals to be achieved by using Core Components are:

- to standardize data across industries and internationally

- to use internationally-maintained registries of data to support re-use, extension and harmonization of data across industries and cultures

- to be independent of syntax, and, at the same time, to enable bindings to commonly used syntaxes like EDIFACT and XML

In summary, the ebXML Framework envisions industry communities collaborating to develop re-usable building blocks that are made available in internationally-managed repositories. Key elements of this vision are to use a combination of naming rules, definitions, business terms, business context and qualifiers to facilitate the process of finding and re-using components in these repositories.

Considering the purpose of a definition, which is to give the meaning of a concept, it is quite clear that definitions are the heart of an idea like Core Components. The components are like concepts, and to find or compare concepts, the meanings or definitions are what have to be compared, as opposed to relying on a comparison of the terms that name the concepts. These can be misleading as discussed below.

Data integration and schema transformation present similar problems. Concepts in the data schemas have to be matched, and it is the definitions, again, that are decisive and not the field names — which are a kind of term to name the concepts — since terms are ambiguous. The same term may designate different concepts in different schemas. Or, different terms may designate the same concept, which is a common problem in cross-industry communication and data integration[2], since each industry has its own jargon, even for business terms. Industry-specific terms for data exchange are commonly documented in an *implementation guideline*. These can be very complex documents as shown in this quote:

> *"Taking xCBL V3.0 as an example, the ORDER allows more that 20,000 elements. It is only due to the implementation guideline that it is possible to specify which elements are actually used in the data exchange of business partners.*"[3]

The process of tailoring standard electronic documents is often multi-level. An industry group defines and documents a "subset" for use in that industry. Then a company in that industry further refines and documents the industry subset, which refinement, in turn, might be fine-tailored for use in a specific process. IT integration consultants can start with these implementation guidelines and add more precise or supplementary definitions, as they build a database of schemas and schema mappings to aid them in future integration projects.

---

[2] Confirmed by Core Components User's Guide, 4.1.16 Conclusion
[3] Quoted from http://www.gefeg.com/en/edifix/xml-spy.htm

The implementation guidelines created in each case, with their definitions and annotations, are the basis for interpreting the exchanged documents and mapping them to in-house formats. Moreover, they can be used to generate program artifacts like XML schemas, database tables, text messages, Web Forms, etc. This section is based mainly on [Un03], [Un04] and [KW02].

2.4 Other uses of terminology management in IT

As the previous discussion has shown, terms and definitions, which are the concern of terminology management, are the key to achieving the Semantic Web, to achieving data-interoperability, as well as to solving some AI and NLP problems. In general, wherever definitions or glossaries are needed in IT, terminology management is applicable. This includes the requirements gathering process, software specification and design, in other words, all the early phases of software development.

3 Basic concepts related to definitions

A number of concepts are important in talking about definitions: term, subject field, intension, extension, ambiguity and vagueness.

3.1 Term

A *term* is a linguistic expression of a single specific concept from a particular subject field for a particular purpose. Purpose is very important as we will see later. The term can take many forms. The next table shows the names of the various forms with examples.

TERM FORM	**EXAMPLE**		**TERM FORM**	**EXAMPLE**
Single word	*machine*		Multi-word	*rear-wheel-drive vehicle*
Set phrase	*Hawks and doves*		Collocation	*file a patent (words can occur separated in text)*
Short form	*m (for meter)*		Boilerplate	*(a set text that describes how to handle a hazardous chemical)*

Table 1 Types of Terms

Terms can be any part of speech: noun, verb, adjective, or adverb. Nouns and verbs are most usual for inclusion in a glossary or terminology collection. Although, verbs might be somewhat rare, since they can be represented by a corresponding nominal form like a gerund, e.g., *swimming* instead of *to swim.*

Examples of terms in software engineering are: database table names, field names, class names, attribute names, data-type names and method or functions names. These are the types of things that are often included in glossaries of software specifications.

In dealing with concepts, terminology collections are similar to ontologies and different from dictionaries. This next table shows how a terminology collection differs from a dictionary or lexicon by showing the differences between the entries in each.

LEXICAL ENTRY	**TERM ENTRY**
Pertains to one word	Pertains to one concept
Lists multiple senses of the word	Lists the terms assigned to the concept
Arranged alphabetically	Arranged in accordance with a concept system, e.g., a taxonomy
Defined based on general knowledge	Defined by careful relation to other concepts in the system
	Facilitates translation, since one concept can be related to the terms for it in multiple languages.

Table 2 Comparison of lexical and term entries, based on [WB97] p.328

As an example, *cardinal*, the noun, has one entry in the American Heritage® dictionary with five senses: a Roman Catholic high church-official, a color, a bird, a cloak and a number. In a terminology collection, which covered all five senses, there would be five entries, one for each word sense.

3.2 Singular and general terms

The characterization of *term* given previously was over-simplified. Logic, which has been concerned with terms and definitions since at least the time of Socrates (5th century B.C.), distinguishes between singular and general terms.[4] General terms are also called universal terms or class terms. A term is singular if its intent is to refer to exactly one thing, one individual. A proper name like *Earnest Worthing* or *Karlsruhe* is singular. A definite description like *the King of France*, is also singular. A general term like *fictional character,* on the other hand, refers to a class of things. It covers many individuals of a certain kind.

[4] This section benefited greatly from the discussion in [KM66], especially chapter 2.

So, to be more accurate the description of *term* should be amended to read: a term is a linguistic expression of a single specific concept (general term) or object (singular term) from a particular subject field for a particular purpose.

It has to be admitted, however, that there are no rules for deciding if a term is singular or general. Abstract terms like *heat* or *justice* are talked about in the singular, e.g., *justice is blind*. But, are they individuals? Similarly terms which are usually considered to be general like *spotted owl* are sometimes used in a singular sense, as in: *The spotted owl is disappearing*. This statement is singular in that it talks about one thing: the species spotted owl. Given these ambiguities, it should not be surprising if there are often disputes about whether something is an individual or not. However, for the purposes of creating Semantic Web ontologies a pragmatic solution might suffice: If all the statements (axioms) about a term with a given meaning are singular, then that term is singular.

Another characterization of terms is to say that they are the basic units of Aristotelian term logic. In the Aristotelian term relation: *All cats meowing in the moonlight are worthy of extermination,* the two terms being related are: *cats meowing in the moonlight* and *things worthy of extermination*. (Example from [KM66] p. 30).

3.3 Intension and extension (both spelled with an 's')

The intension of a concept is its set of distinguishing characteristics, and is essentially a formula for testing if the concept is applicable to something. The extension of a concept, on the other hand, is determined by its intension, and is the set of all those classes and objects that each have the distinguishing characteristics of the concept.

An example, using the word *cardinal,* should make this clearer. Sense number one of *cardinal* is defined by the American Heritage ® as follows:

> **"1.** *abbr*. **Card.** *Roman Catholic Church* A high church official, ranking just below the pope, who has been appointed by a pope to membership in the College of Cardinals."

The subject field here is the *Roman Catholic Church*, and the characteristics (intension) of *cardinal* are: is a high church-official; ranks just below the pope; and is appointed by the pope to membership in the College of Cardinals. Tracing the concept system further, would lead to the characteristics of the concept, *the College of Cardinals*: elects new popes; and advises the pope. Intension determines extension, so that the extension of *cardinal* is the set of all members of the College of Cardinals. As seen in this example a characteristic can be just about anything that can be said about something: an attribute, relationships with other concepts, or the function or behavior associated with the concept.

The standard form of definition is one by intension, that is, it lists the distinguishing characteristics of the concept. However, as will be seen later, some kinds of extensional definition are possible too.

A term can be meaningful and yet have a null extension, e.g., *the present king of France* is a singular term that currently has a null extension. Extension and existence are not exactly the same thing. Fafnir, the Norse dragon who guards the Nibelung treasure, can be considered part of the extension of mythical dragons, whether he exists or not.

There is much debate about what is real. Nominalism, as espoused by Berkeley and Hume in the eighteenth century, says only individuals are real, i.e., reality consists entirely of individual things, with nonphysical things allowed. General terms are only names for groups of individuals, hence the name for this position, *nominalism*. Realism, represented by Plato, says that only general or class terms are 'real'. The realm of coming to be and passing away, the individual, can't be 'real' in his sense, since for Plato only the permanent is the 'real'.

3.4 Ambiguity and vagueness

There are at least two different ways in which the meaning of a given term can be unclear. It can be ambiguous or vague. Most terms are both. Ambiguity is caused by the fact that most terms have more than one meaning. On average an English word has about seven different meanings or senses. Many have more. A list of terms compiled from various ASTM (American Society for Testing) committees contained thirty different definitions for the word *lot*. [WB97] p. 199.

Vagueness is a feature of concepts that remains even after ambiguity has been resolved by deciding which single sense of a term is the intended one. Vagueness means that it is not always possible to decide in every case whether a concept applies or not. Sense one of *city* from the American Heritage® is an example.

> "**1.** A center of population, commerce, and culture; a town of significant size and importance."

This concept just doesn't have enough precision to decide in every case if something is a city or not.

A very good short article on vagueness in [Ed72] suggests that if a definition is to capture the essence of a word, it should be as vague as the word itself. For example *adolescence* can be defined as *the period of life between childhood and adulthood*. *Childhood* and *adulthood* have fuzzy boundaries that match the fuzzy boundaries of *adolescence*. The same article points out that vagueness is often even desired and that achievement of its opposite, total precision, is a *chimera*, an impossible dream. Politics is an arena in which vagueness can be used to keep options open. A famous example is the recent U.N. Security Council Resolution 1441, which says that Iraq would face "serious consequences" for failing to comply. This phrase specifically avoided traditional terminology, "all available means", which looks just as vague, but whose meaning at the U.N. is not ambiguous.

The fact that the meaning of terms changes over time is yet another complicating factor that makes things even more challenging. The term *heat*, discussed later, is an example of a term whose meaning has changed significantly.

4 Terminology management

The introduction highlighted the use of terminology management to solve interoperability problems at the data level. At the business level it is also of great importance, for reasons of efficiency, cost, quality and safety. By promoting the use of the same terms for the same concepts, terminology management helps to avoid misunderstandings that waste time and money, and can negatively affect quality. Avoidable misunderstandings related to product liability and environmental regulations can even result in law suits and financial losses. Worse yet misunderstanding of terms can result in loss of life. As related in [WB97] p. 219, a misinterpretation of the term *glycol* in a recipe for medicinal syrup resulted in the deaths of many children.

The benefits of terminology management and the usual process for creating terminologies are directly applicable to IT projects. The first step is to determine the boundaries of the subject field, and to locate experts in that field. The next step is to work with the domain experts to gather the artifacts that are the source of the terms and concepts. These can range from interviews or spoken discourse to handbooks, textbooks, standards, authoritative texts, laws, regulations and so on. Terms, concepts and initial definitions are selected and extracted from these artifacts. In the next step the collected concepts are organized into a concept system, with class and part hierarchies being the most common systems used, although as will be seen later, other systems may sometimes be appropriate. Finally, in an iterative process the concept definitions and the concept system are refined to create the terminology collection.

During the process of creating the terminology collection, information about each concept and its associated terms is recorded: terms that denote the concept; including short forms; the preferred term for the concept; suggested terms for the concept; used but deprecated terms for the concept; unequivocal (*monosemous*, meaning *single meaning*) terms for the concept; equivocal (*polysemous*, meaning *many meaning*s) terms for the concept; any standard symbols associated with the concept; and, finally, a classification of each term as being specific to the subject field, borrowed from a related subject field or taken from general language.

Terminology entries should also contain uses of the term in context. *Context* in terminology means something different than in IT. It means the section of text in which a term is used, e.g., a sentence, paragraph or longer extract. If a good definition of a concept already exists, that should also be included with a full bibliographic reference to its source.

Most importantly, the entry must name the subject field to which the terms belong, especially when terms have multiple meanings, which is often the case. Otherwise, when someone looks up a term in the collection and gets back multiple entries, it will be very difficult to decide which entry is the desired one.

5 Definitions: some examples

The previous section outlined a method for creating a collection of concept definitions. But, there are no recipes or algorithms for writing good definitions. The only reliable test of a definition is in the understanding of the reader. For this reason the definition has to be written with the reader's assumed prior knowledge and cognitive purposes in mind. The importance of purpose cannot be over-emphasized, and will be highlighted in the example definitions that follow. The types of definitions discussed are intensional, extensional, ostensive, theoretical, operational and ad-hoc.

5.1 Definition by genus and difference; and the importance of purpose

The classic way to define a concept is by genus and difference. In logic a *genus* is a super-class of the sub-class to be defined, and this sub-class is called the *species*. The next table lists quasi-synonymous terms from IT, linguistics and logic, which are used to talk about super-classes and sub-classes. As used in logic *genus* and *species* are relative terms, as are the other terms given in the table. Using figure 2 as an example, taken by itself *motorized-vehicle* is neither a species nor a genus. It is a species relative to *vehicle* and a genus relative to *car*.

HIGHER LEVEL	LOWER LEVEL	SAME LEVEL
Super-class	Sub-class	-
Superordinate	Subordinate	Coordinate
Broader	Narrower	-
Hypernym	Hyponym	Co-hyponym
Generalization	Specialization	-
Genus	Species	Coordinate species

Table 3 Term correspondence (all relative terms)

The following definitions exemplify two things: the importance of purpose and an example of a classical definition by genus and difference. The simple concept system underlying the example definitions is shown in the next figure.

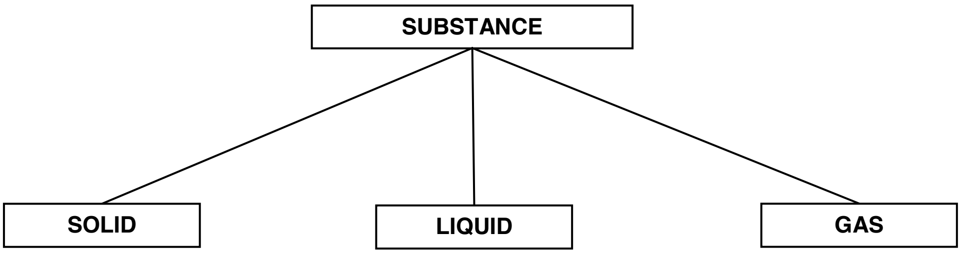

Figure 1: Simple concept system, a taxonomy

The intension of a concept, that is, its distinguishing characteristics, depends on purposes and points of view as the following definitions quoted from [ISO87] show. The first defines *liquid* for the purposes of hydromechanics, and the second for the purposes of thermodynamics.

> Subject-field hydromechanics: “a liquid is a substance which is incompressible, very dense and capable of flowing”.

The intension or set of characteristics are: is a kind of substance; is incompressible; is very dense; is capable of flowing.

> Subject-field thermodynamics: “a liquid is a substance in a condensed state, intermediate between a solid and a gas”

The first definition is an example of the pure classical method of definition by genus and difference. The genus is *substance*, that is to say, a liquid is a kind of substance. The rest of the definition lists the hydromechanical characteristics that **differentiate** liquid from other kinds of substances. Definition by genus and difference relies on the idea that a species possesses all the characteristics of its genus plus at least one additional characteristic that differentiates it from the other species of its genus.

The second definition is not quite so classical, since it essentially defines liquid by saying that it’s not a solid and not a gas; it’s something in between. This is a technique that can work if the number of other species of a genus is small, as it is here, comprising as it does only solid, liquid and gas.

Just as a reminder, characteristics can be just about anything that can be said about something. In the next example from [Co68] p. 115, the origin is the distinguishing characteristic of *Stradivarius*.

> A *Stradivarius* is a violin which was made in the Cremona workshop of Antonio Stradivari.

It is difficult to pick out the distinguishing characteristics, that is, the intension, of a concept as the following story from [Co68] p. 116 illustrates. Not long after the days of Plato, the philosophers in the Academy at Athens were seeking a definition of *man*. After heated debate they agreed to define *man* as *featherless biped.* Shortly thereafter, Diogenes, the cynic, made mischief by throwing over the wall into their midst a plucked chicken.

5.2 Necessity and sufficiency

Necessity and sufficiency provide another way to think of intension and throw some light on the previous academic definition of *man*. Necessary and sufficient characteristics (or conditions) are related to the super-class/sub-class relation as follows. Characteristics that are necessary to a class define a super-class of that class. For example, *featherless biped* is a super-class of *man*. By contrast, characteristics that are sufficient to a class define a sub-class of that class. *Investing 150,000 pounds in the country* is sufficient to *qualify for residency in the UK*, that is, people who qualify under this rule are a sub-class of people who qualify for residency.

It is the set of characteristics that is both necessary and sufficient that defines a class by intension. In the previous definition of *man,* the characteristics were only necessary, but not sufficient. Rather than defining *man,* these necessary conditions only defined a super-class of *man*, a super-class that also included plucked chickens.

5.3 Theoretical definition

A theoretical definition embodies a theory, and will only be accepted if this theory is also accepted. An example from the Oxford English Dictionary (OED), 2nd Edition is:

> *heat* is "a form of ENERGY, viz. the kinetic and potential energy of the invisible molecules of bodies"

Acceptance of this definition requires "belief" in molecules and the different types of energy. However, not that long ago, there was another definition of heat that had its place in a radically different theory:

> *heat* is "an elastic material fluid of extreme subtlety attracted and absorbed by all bodies" (OED, same entry as previous definition)

As Thomas Kuhn argues in his book [Ku70], and these definitions of heat confirm, science advances not so much by accumulation as by revolutions that throw out the old sets of definitions and replace them by new sets of definitions. After such a revolution a new scientific consensus forms around new assumptions, values, theories, corresponding definitions and ways of doing science. This new constellation of theory and practice is known as a *paradigm*, a word coined and defined by Thomas Kuhn, which has subsequently become something of a buzzword in other fields. Einstein's definition of *simultaneity* is an example (mentioned in [Ay52] p. 152) of a definition that revolutionized physics. Similarly, in mathematics the Cauchy-Weierstrauss definition of limit and continuity paved the way for whole new branches of mathematics according to [Ch02] page Roman numeral x.

It should also be observed here that all definitions embody an element of theory; even classification as shown by the fact that there are three methods of classifying living organisms: cladistics, phenetics and evolutionary systematics. Each classification system has its own distinct purposes and associated theories.

5.4 Example of complex multi-dimensional concept system and a definition

This example taken from [WB97] shows that concept systems are often multi-dimensional. Using this concept system as a guide, the concept *car* can be defined as follows: a *car* is a motorized vehicle for transporting passengers by land

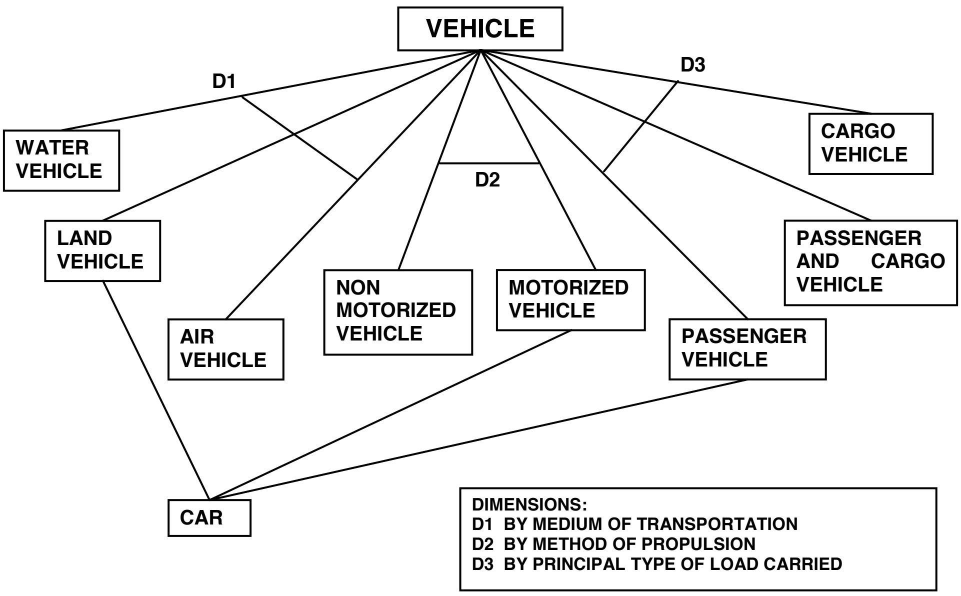

Figure 2: Multi-dimensional concept system based on [WB97] p. 135

5.5 Ad-hoc definition based on a practical concept system

This example from [WB97] shows how purpose guides the choice of concept system and definition. The purpose is to instruct an operator about a machine. Definitions are the preliminary to that, and are based on a very simple concept system, which is all that is necessary for this purpose.

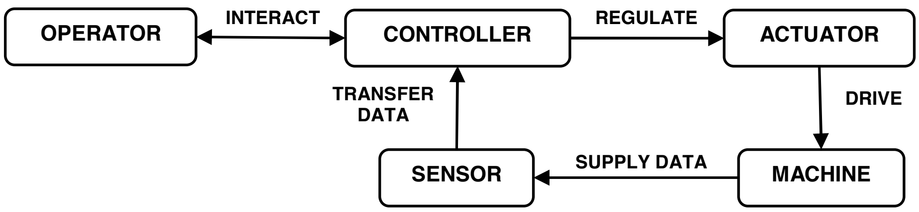

Figure 3: Simple concept system for machine operator based on [WB97] p. 446-447

Definitions can practically be "read off" the diagram of the concept system. Looking at the simple diagram given, an *actuator* can be defined as follows: an *actuator* is the device that is regulated by the controller, and which drives the machine. This definition is much more useful to an operator than the standard definition from IEC 1992, quoted by [WB97], which is:

> "actuator (electric) An electric transducer that converts an electrical signal into a signal of any kind, such as mechanical displacement."

The definition of verbs can also be "read off" the concept system: For example, *to regulate* an actuator is to adjust its state based on feedback from sensors and a dialogue with the operator who evaluates the sensor data.

5.6 Definition of types of terms: verbs, adjectives, correlatives

The type of term influences the form of the definition. As shown in the previous section a verb can be defined by relating it to other verbs, possibly a synonym or a troponym (as defined by WordNet).

For an adjective it is often useful to give opposites or even to list an entire range of adjectives, e.g., *dark* can be defined as *lacking or having very little light.*

Special attention should be given to correlatives, not only in definitions, but also in data modeling. The Oxford English Dictionary, 2nd Edition (OED) defines *correlative* as "having a reciprocal relation such that each necessarily implies, or is complementary to, the other" Examples are: *father and son*, *husband and wife*, *employee and employer*, *slave and master* or *owner and owned object*. In general, it is a good idea to mention the complement of a correlative when defining it. For example, an *employee* is *someone who is employed by an employer, usually working for a wage in return for his or her services*.

When modeling a correlative for IT purposes, it is most flexible to model it as an explicit relation between the two types of related objects, e.g., a relation between persons and employers, where employers can be either persons or companies. This relation could, for example, be a separate database table, each of whose records would relate one person to one employer. This modeling choice makes it possible to easily do things that would be difficult or impossible to do if an employee were modeled as a class instead. For example, an explicit model of the relation makes it easy to record histories of employment by adding the period of employment as a field in the table. It also makes it possible for one person to have multiple employment relations at the same time. These features are particularly important in a world of so-called virtual organizations, where one person may work successively for related short-lived companies or may even work for multiple companies at the same time. Older class models of employee are inadequate to provide this kind of flexibility. Some of these older models don't even include the name of the employer. They just assume it is the one company that owns the database.

5.7 Operational definition

"An operational definition is a procedure agreed upon for translation of a concept into measurement of some kind." - W. Edwards Deming. Operational definitions are extremely important in business, where they are the basis for quality control and determination of conformance to business agreements.

Here is an example of why it is important to have operational definitions of concepts like *flat*.

> "On October 15, 1970, the West Gate Bridge in Melbourne, Australia collapsed, killing 35 construction workers. The subsequent enquiry found that the failure arose because engineers had specified the supply of a quantity of flat steel plate. The word *flat* here lacks operational definition so there is no basis for accepting or rejecting a particular shipment or for controlling quality."

The above quoted material is from:
http://encyclopedia.thefreedictionary.com/operational%20definition

5.8 Definition by example or extension

Terms can be defined by extension or example, which is listing the species or individuals to which the concept is applicable, for example, *the planets of the solar system* are Mercury, Venus, Mars, etc. This technique can only work if this set is small and finite. But, in general giving examples has two main drawbacks. First, any individual can be an example of many things, e.g., *Erasmus* is an example of a man, a renaissance humanist, etc. Second, there is no guidance given to determine if a new object belongs to the set or not.

Ostensive definition is definition by pointing. But, it has the drawback of often being unclear as to what exactly is being pointed at. However, a quasi-ostensive definition, e.g., pointing to an image of a chair and saying “this type of furniture”, is a workable technique, particularly in the age of multi-media, where text, image and sound can be combined. This example of a quasi-ostensive definition is from [Co68] p. 109.

5.9 Universal concepts are difficult to define

It is very difficult to define universal concepts like object, concept, knowledge, characteristic, quality, etc. They cannot be defined using genus and difference since there is no genus for them. In the OED, 2nd Edition the definitions of *characteristic*, *attribute, property* and *quality* pretty much go around in a circle, using each other in their definitions.

6 Summary

It is hoped that the examples and discussions in this paper have sensitized the reader to the importance of words and definitions, as well as the associated difficulties and subtleties. Possibly, the reader may now agree with Erasmus when he says:

“In principle, knowledge seems to be of two kinds, of things and of words. Knowledge of words comes earlier, but that of things is the more important. But some, the ‘uninitiated’ as the saying goes, while they hurry on to learn about things, neglect a concern for language and, striving after a false economy, incur a very heavy loss. For since things are learnt only by the sounds we attach to them, a person who is not skilled in the force of language is, of necessity, short-sighted, deluded, and unbalanced in his judgment of things as well. Finally you may observe that none are more given to constant quibbling over the minutiae of language than those who boast that they pass over mere words and concentrate on the matter itself.”[5]

Might more concern with “mere words” decrease arguments about designs, facilitate communication with users, and thereby result in a speed-up of the whole software development process?

7 Outlook and future work

This was just a first attempt to pull together some ideas. The next step is to apply the ideas in a software engineering context. Feedback from anyone else who attempts to do this would be highly welcome. A future paper would also discuss some of the common errors of definition like circularity. An example of a circular definition is: A *business process step* is *a step in a business process*.

---

[5] Quoted from [Da03], Vol. 24 *On the Method of Study*

8 Acknowledgements

This paper is partly funded by the European Commission through the Athena Integrated Project (www.athena-ip.org). It does not represent the view of the Commission, and the author is solely responsible for the paper's content.

9 Literature

[Ay52] Ayer, Jules Alfred: Language Truth and Logic. Dover, New York, 1952.
[Ch02] Chahal, J. S.: Historical Perspective of Mathematics. Unpublished manuscript, 2002
[Co68] Copi, Irving M.: Introduction to Logic, 3rd Edition. Macmillan, 1968.
[Da03] Dalzell, Alexander et al: Collected Works of Erasmus. Univ. of Toronto Press, 2003.
[Ed72] Edwards, Paul et al: The Encyclopedia of Philosophy. Macmillan, 1972.
[Fe98] Fellbaum, Christiane et al: WordNet: an electronic lexical database. MIT Press, 1998.
[Fr04] Frankel, David; Hayes, Pat; Kendall, Elisa; McGuinness, Deborah: MDA Journal July 2004, A BPT column from www.bptrends.com.
[Gä00] Gärdenfors, Peter: Conceptual Spaces : the geometry of thought. MIT Press, 2000.
[ISO87] International Organization for Standardization: Principles and methods of terminology. ISO 704. ISO, 1987.
[Jo86] Jones, Karen Spark: Synonymy and Semantic Classification. Edinburgh Univ. Press, 1986.
[KM66] Kupperman, Joel; McGrade, Arthur S.: Fundamentals of Logic. Doubleday, 1966.
[Ku70] Kuhn, Thomas S.: The Structure of Scientific Revolutions, 2nd Edition. The University of Chicago Press, 1970.
[KW02] Kotok, Alan; Webber, David R. R.: ebXML The New Global Standard for Doing Business over the Internet. New Riders Publishing, 2002.
[Un03] UN/CEFACT. Core Components Technical Specification – Part 8 of the ebXML Framework. 15 November 2003, Version 2.01
[Un04] UN/CEFACT. Core Components User's Guide. 13 March 2004
[Us98] Uschold, Mike; King, Martin; Moralee, Stuart; Zorgios, Yannis: The Enterprise Ontology. in The Knowledge Engineering Review, Vol. 13, Special Issue on Putting Ontologies to Use. 1998.
[WB97] Wright, Sue Ellen; Budin, Gerhard: Handbook of Terminology Management Volume 1. B.V. Publishing, 1997.