The role of terminology management in creating ontologies

Susan Thomas, SAP Research, SAP AG, Karlsruhe Germany
based on a paper published in BTW2005: "The importance of being earnest about definitions"

1 Motivation

The purpose of this talk is to explain what terminology management is all about and to show how it can aid the creation, maintenance and re-use of ontologies. One of the central tasks of terminology management is writing good definitions of terms, so that this talk will explain the key concepts related to definitions, and show how these concepts relate to OWL ontologies. Examples are used to illustrate all major points.

It is common to draw an analogy between ontologies and definitions of terms. For example, the authors of [Fr04] 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", that is to say term definitions. In the context of the semantic web these "definitions" are expressed formally using ontologies. But such formal ontologies should always be associated with natural language definitions of the terms in the ontology. 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, either for the purpose of making changes or to decide if it is suitable for use in a new context. Terminology management, which is explained in this talk, provides a methodical way to start the development of ontologies, and to write definitions.

In spite of the cited advantages of having natural-language definitions, there may still be some who imagine that the formal statements are enough. But, the experience of the developers of WordNet shows that this view is mistaken.

WordNet is a hand-crafted electronic lexical database of English. Its coverage is 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. This makes it similar to a simple ontology.

Meanings or concepts in WordNet are represented by sets of words with essentially the same meaning, that is, synonyms. These sets are called synsets. Initially WordNet contained no definitions of concepts/synsets at all. Definitions were considered to be superfluous because the creators of WordNet thought that the meanings of words could be easily inferred from the semantic and lexical relations captured in the database. 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] p. Roman Numeral xx, they learned the hard way the importance of definitions.

2 Basic concepts related to definitions

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

2.1 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 is shown in the talk. A term can take many forms, such as single word, multi-word, collocation, set phrase, short form or boilerplate.

2.2 Intension and extension

The standard form of definition is by intension. The intension of a concept is its set of distinguishing characteristics. 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 follows. Sense number one of cardinal is defined by the American Heritage ® as:

"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*. 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.

3 Terminology management – based on textbook [WB97]
Terminology management is of great importance in business 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 to corporations.

The benefits of terminology management and the usual process for creating terminologies are directly applicable to ontology creation. 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 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.

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 terminology entry must name the subject field to which the term belongs, 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.

4 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 given. The talk will also include a few examples of inadequate definitions to show that it is not easy to write good definitions.

4.1 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 class to be defined which is called the species. The examples given show two things: the importance of purpose and a definition by genus and difference.

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.

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.

4.2 Necessity and sufficiency, related to primitive and defined classes in OWL

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 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. The OWL concepts of primitive and defined class can be similarly explained.

4.3 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)

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 Acknowledgements

6 Literature

[Co68] Copi, Irving M.: Introduction to Logic, 3rd Edition. Macmillan, 1968.
[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.
[ISO87]International Organization for Standardization: Principles and methods of terminology. ISO 704. ISO, 1987.
[WB97]Wright, Sue Ellen; Budin, Gerhard: Handbook of Terminology Management Volume 1. B.V. Publishing, 1997.