What are the differences between a vocabulary, a taxonomy, a thesaurus, an ontology, and a meta-model?

(This excellent overview was written by Woody Pidcock of the Boeing company and posted at `metamodel.com`. It has been edited slightly so it could be archived here.)

I will answer this question one step at a time. To keep this answer focused on the question, I will use other concepts that I will not define here.

A **controlled vocabulary** is a list of terms that have been enumerated explicitly. This list is controlled by and is available from a controlled vocabulary registration authority. All terms in a controlled vocabulary should have an unambiguous, non-redundant definition. This is a design goal that may not be true in practice. It depends on how strict the controlled vocabulary registration authority is regarding registration of terms into a controlled vocabulary. At a minimum, the following two rules should be enforced:

1. If the same term is commonly used to mean different concepts in different contexts, then its name is explicitly qualified to resolve this ambiguity.
2. If multiple terms are used to mean the same thing, one of the terms is identified as the preferred term in the controlled vocabulary and the other terms are listed as synonyms or aliases.

A **taxonomy** is a collection of controlled vocabulary terms organized into a hierarchical structure. Each term in a taxonomy is in one or more parent-child relationships to other terms in the taxonomy. There may be different types of parent-child relationships in a taxonomy (e.g., whole-part, genus-species, type-instance), but good practice limits all parent-child relationships to a single parent to be of the same type. Some taxonomies allow poly-hierarchy, which means that a term can have multiple parents. This means that if a term appears in multiple places in a taxonomy, then it is the same term. Specifically, if a term has children in one place in a taxonomy, then it has the same children in every other place where it appears.

A **thesaurus** is a networked collection of controlled vocabulary terms. This means that a thesaurus uses associative relationships in addition to parent-child relationships. The expressiveness of the associative relationships in a thesaurus vary and can be as simple as â€œrelated to termâ€ as in term A is related to term B.

People use the word **ontology** to mean different things, e.g. glossaries & data dictionaries, thesauri & taxonomies, schemas & data models, and formal ontologies & inference. A formal ontology is a controlled vocabulary expressed in an ontology representation language. This language has a grammar for using vocabulary terms to express something meaningful within a specified domain of interest. The grammar contains formal constraints (e.g., specifies what it means to be a well-formed statement, assertion, query, etc.) on how terms in the ontologyâ€™s controlled vocabulary can be used together.

People make commitments to use a specific controlled vocabulary or ontology for a domain of interest. Enforcement of an ontologyâ€™s grammar may be rigorous or lax. Frequently, the grammar for a "light-weight" ontology is not completely specified, i.e., it has implicit rules that are not explicitly documented.

A **meta-model** is an explicit model of the constructs and rules needed to build specific models within a domain of interest. A valid meta-model is an ontology, but not all ontologies are modeled explicitly as meta-models. A meta-model can be viewed from three different perspectives:

1. as a set of building blocks and rules used to build models
2. as a model of a domain of interest, and
3. as an instance of another model.

When comparing meta-models to ontologies, we are talking about meta-models as models (perspective 2).

Note: Meta-modeling as a domain of interest can have its own ontology. For example, the CDIF

Family of Standards, which contains the CDIF Meta-meta-model along with rules for modeling and extensibility and transfer format, is such an ontology. When modelers use a modeling tool to construct models, they are making a commitment to use the ontology implemented in the modeling tool. This model making ontology is usually called a meta-model, with â€œmodel makingâ€ as its domain of interest.

Bottom line: Taxonomies and Thesauri may relate terms in a controlled vocabulary via parent-child and associative relationships, but do not contain explicit grammar rules to constrain how to use controlled vocabulary terms to express (model) something meaningful within a domain of interest. A meta-model is an ontology used by modelers. People make commitments to use a specific controlled vocabulary or ontology for a domain of interest.

Additions

Michael Uschold of the Boeing company has provided some excellent additions:

There is some good material here, and it is a very difficult question to answer well. A good way to get clear is to say first what they all have in common, and then look at the things that some have that the others do not.

What controlled vocabularies, taxonomies, thesauri, ontologies, and meta-models **all have in common** are:

- They are approaches to help structure, classify, model, and or represent the concepts and relationships pertaining to some subject matter of interest to some community.
- They are intended to enable a community to come to agreement and to commit to use the same terms in the same way.
- There is a set of terms that some community agrees to use to refer to these concepts and relationships.
- The meaning of the terms is specified in some way and to some degree.
- They are fuzzy, ill-defined notions used in many different ways by different individuals and communities.

The **major differences** that distinguish these approaches:

- How much meaning is specified for each term?
- What notation or language is used to specify the meaning?
- What is the thing for? Taxonomies, thesauri, ontologies, and meta-models have different but overlapping uses.

A **controlled vocabulary** may have no meaning specified (it could be just a set of terms that people agree to use, and their meaning is understood), or it may have very detailed definitions for each term.

A **taxonomy** has additional meaning specified via whatever the meaning of the hierarchical link is. In a traditional 'taxonomy' the meaning is generalization/specialization or 'is a kind of', depending on what direction you are going. These days the word 'taxonomy' is used to refer to other kinds of hierarchies with different meanings for the links (e.g., part of, broader topic than, instance of). Sloppy taxonomies will not identify explicitly what the meaning of the link is, and there may be different meanings. If a taxonomy has a variety of very carefully defined meanings for the hierarchical link, then it bears a stronger resemblance to an ontology.

A **thesaurus** has two kinds of links: broader/narrower term, which is much like the generalization/specialization link, but may include a variety of others (just like a taxonomy). In fact, the broader/narrower links of a thesaurus is not really different from a taxonomy, as described above. A thesaurus has another kind of link, which typically will not be a hierarchical relation, although it could be. This link may not have any explicit meaning at all, other than that there is some relationship between the two terms.

The word **'ontology'** has been used to refer to all of the above things. When used in the AI/Knowledge Representation community, it tends to refer to things that have a rich and formal logic-based language for specifying meaning of the terms. Both a thesaurus and a taxonomy can be seen as having a simple language that could be given a grammar, although this is not normally done. Usually they are not formal, in the sense that there is no formal semantics given

for the language. However, one can create a model in UML and a model in some formal ontology language and they can have identical meaning. It is thus not useful to say one is an ontology and the other is not because one lacks a formal semantics. The truth is there is a fuzzy line connecting these things.

Look here for a paper describing the many uses of ontologies: A Framework for Understanding and Classifying Ontology Applications (http://citeseerx.ist.psu.edu/viewdoc /download?doi=10.1.1.39.6456&rep=rep1&type=pdf Rob Jasper and Mike Uschold.)

There is a very close relationship between a meta-model and an ontology, but it is not necessarily equivalence.

IF: you create an ontology, which is a set of terms naming concepts (classes) and relations, and you use that vocabulary to create a set of data (instances of the classes, and assertions that the instances are related to each other according to the specific relations in the vocabulary), and you think of the set of data you create as the model of your domain,

THEN: the ontology is the meta-model and the set of data created is the model.

In this case, there is little if any useful distinction to be drawn between an ontology and a meta-model. However, meta-models aren't always used in this way to connect to specific models, which is one of the primary uses of ontologies.

See also: "Where are the Semantics on the Semantic Web?" by Mike Uschold. This paper is an attempt to clarify some issues about different views of semantics and how they apply to the Web. It introduces the idea of a semantic continuum explaining the differences between vocabularies, taxonomies, thesauri, etc. To appear in AI Magazine sometime soon. See http://lsdis.cs.uga.edu/events/Uschold-talk.htm for abstract, paper, and presentation.

Last modified on 08/04/10 19:38:48