CLARITY IN THE USAGE OF THE TERMS ONTOLOGY, TAXONOMY AND CLASSIFICATION

Reinout van Rees
Civil engineering informatics, Delft University of Technology
reinout@vanrees.org

SUMMARY

There is a lack of clarity when discussing the following three terms: classifications, taxonomies and ontologies. A general cause of confusion is caused by a trend, observed at a recent conference, to use the most fashionable of the three terms: "ontology", without further qualifications. This lack of clarity prompted the writing of this paper with the aim of clarifying the terminology used. A detailed extract from all relevant papers of the EBEW-conference 2001 on the use of the three terms was made to provide a quantification of the usage of the three terms. The recommendation by the author is to make a specific choice of label ("ontology", "taxonomy" or "classification") for your dataset and to provide further qualification on top of that label to remove ambiguity.

INTRODUCTION

There is a lack of clarity when discussing the following three terms: classifications, taxonomies and ontologies. A general cause of confusion is caused by a trend, observed at a recent conference, to use the most fashionable of the three terms, "ontology", without much qualification. This lack of clarity prompted the writing of this paper with the aim of clarifying, or at least discussing, the terminology used. The goal is not to discuss available building and construction ontologies et cetera, but to promote a clearer use of the terms (especially "ontology") in building and construction research.

It is almost impossible to define one of these three terms in a clear way as their incarnations almost invariable incorporate functionality found in one of the others' definitions. There is almost always a mix between two or three of the terms. An example is the UN Standard for Product and Services Classification (UNSPSC, http://www.unspsc.org) classification. It consists of a unique number for each product(category) and a label. But UNSPSC also adds a little explanation, which makes it a bit ontology-like. And it's got a hierarchy, which makes it taxonomy-like.

The definitions of the three terms are the necessary beginning of the paper. This is followed by a discussion of the differences between the terms. An extract from all relevant papers at a recent conference will provide further input on the use and definition of the terms. New definitions are provided, followed by a suggestion on how to use this terminology.

FIRST DEFINITIONS

To provide a starting point, the Merriam-Webster (http://www.m-w.com) dictionary's entry for "ontology", "taxonomy" and "classification" are provided below, coupled with an additional explanation by the author. For the term "ontology" additional definitions are discussed.

Classification

Merriam-Webster definition: *systematic arrangement in groups or categories according to established criteria.*
An example of a classification would be the division of all animals in the classes tasty, edible and not edible.
Most construction specification systems fall into this category, classifying for instance in administrative items or technical items such as HVAC, doors and windows et cetera.

Taxonomy

Merriam-Webster definition: *orderly classification of plants and animals according to their presumed natural relationships.*

A clear example of a taxonomy is the animal kingdom taxonomy. Kingdom "animals", class "mammals", order "carnivores", genus "canis", species "canis lupus", which is the common gray wolf[1]. Other members of the genus "canis" are the dog and the jackal. This is a taxonomy based on the presumed "is a kind of" relation.

A taxonomy can thus best be described as a hierarchy created according to data internal to the items in that hierarchy.

Ontology

Merriam-Webster definition: *a branch of metaphysics concerned with the nature and relations of being or a particular theory about the nature of being or the kinds of existents.*

This is the abstract philosophical notion of "ontology", a more applicable term for this field is "formal ontology" [McGuinness 2002]. [Gruber 1993] (widely cited) provides the definition "a specification of a conceptualisation". An ontology thus provides a set of concepts from a certain domain that are well-specified.

"Ontology" is *the* term used on the internet when discussing the semantic web. The WebOntology working group at W3C emphasises that ontologies (in their definition) are *a machine-readable set of definitions that create a taxonomy of classes and subclasses and relationships between them.*

[McGuinness 2002] states that the minimum requirements of an ontology are a finite set of unambiguously identifiable classes and relationships, including strict hierarchical subclass relationships. Typical, but not mandatory is property specification on class basis.

The DAML [Hendler et al. 2000] working group (also a semantic web technology) almost equates ontology with knowledge base. Also the WebOntology working groups charter talks about a knowledge representation language. Their idea is that a lot of knowledge can be captured as data. A contractor, for instance, could add his in-house knowledge on pile driving to a generic definition of piles, like manpower needed, average profit, et cetera.

Well-specified relationships could provide the building industry with partial solutions for known problems as the fire-resistance of doors: you cannot attribute fire-resistance to a single part of a doorset. You need a specific, certified combination of frame, door, hardware, etc. to obtain the required fire-resistance. These interdependencies cannot be expressed directly in simple object-and-property languages like bcxml [van Rees et al. 2002], [Tolman et al. 2002], but the use of a full-blown ontology (which is relatively easy to do for bcxml) *does* support the expression of these interdependencies.

DIFFERENCES

To gain a clearer understanding of the individual definitions, the three terms are offset against each other. This way, the differences and similarities become more pronounced. The definitions provided above are taken as a basis for the comparison.

Differences between a taxonomy and a classification

The difference between a classification and a taxonomy is that a taxonomy classifies in a structure according to some relation between the entities (see above) and that a classification uses more arbitrary (or external) grounds. As an example of *internal* grounds, spinach is a vegetable and not every vegetable is spinach, so spinach is a subclass of vegetable. The decision to place spinach in the category vegetable is based upon data inherent to the entities, so this would be a piece of taxonomy (a taxonomy with a subclass hierarchy).

---

[1] I've left out the pylum "chordata" and the family "canidae" to make the example clearer. They're here for completeness.

An *external* reason could be for instance classification of building components according to the branches of the building industry. This would lead to a classification, not a taxonomy. A taxonomic relation is a relation between entities in the taxonomy (a subclass relation for instance), a classification relates the entities to something that is external (like branches of an industry or safety classes).

Differences between a taxonomy and an ontology

It has already been mentioned earlier that an ontology resembles both a kind of taxonomy-plus-definitions and a kind of knowledge representation language. Knowledge should not be seen as really "active" artificial-intelligence-type knowledge. Read it as "a lot of information", especially relationships. Often, an ontology will *contain* a subclass-based taxonomic hierarchy. As extra properties can be added to the taxonomy as a definition (and proof) of the chosen hierarchy and as ontologies can contain taxonomic relations, the distinction between an ontology and a taxonomy is often blurred. [McGuinness 2002] uses "taxonomy" interchangeably with "simple ontology".
Adding quantifications to the plain terms "ontology" and "taxonomy" is a good way to obtain clarity. An "ontology with a subclass-based taxonomic hierarchy" leaves less room for doubt than using just the term "ontology".

Differences between a classification and an ontology

The fundamental difference between a classification and an ontology is in the richness of information available. Both provide a list or structure of concepts or classification items. But a classification basically stops at that point. It provides boxes with labels into which to put your items. An ontology provides you with a lot of information about the concepts, including their relationships.
If you classify your information in a classification, you place your data in labelled boxes. If you classify (I use the verb for both) your information in an ontology, you automatically enrich your data with all the information stored in the ontology.

Thesaurus

A term not yet discussed is "thesaurus". In principle, a thesaurus deals only with words, alternatives for those words, synonyms, translations, et cetera. This textual kind of information can be used by (or added to) a classification, a taxonomy and an ontology. For instance, a pure thesaurus (like http://irc.nrc-cnrc.gc.ca/thesaurus) could be enhanced to an ontology, providing both the already available rich text information and formal definitions and properties.

Conclusions

- The core property of a taxonomy is that it possesses a hierarchical structure.
- A taxonomy classifies according to properties internal to the data, a classification can be made according to external criteria.
- Taxonomies tend to mix with simple ontologies. Using more specific terms than just "ontology" (like "ontology with inheritance hierarchy") helps to give more clarity.
- Once a lot of properties and relationships are added to a hierarchical structure, the term "ontology" is better suited than "taxonomy".
- A classification tells you in which box your piece of data is, an ontology tells you *what* your data is.

EXTRACT FROM EBEW 2001

A detailed extract from the relevant papers of the e-business and e-work conference in October 2001 was made. The papers were searched[2] for the use of the terms "classification", "taxonomy" and "ontology". This makes for an very interesting comparison between the three terms and their usage.

---

[2]An automated text search in all pdf files.

Classification

Classification as a verb, not as a noun
[Fersoe et al. 2001], [van Schoubroeck et al. 2001], [Bönke et al. 2001] and [Rutten et al. 2001] see classification as a process, putting things in the most-fitting hole. In these references, the holes are provided by either a taxonomy or an ontology. It's the process that matters, where it's stored is almost secondary.

Classification as a set of categories
This is the most common interpretation [Peters et al. 2001] [Holtkamp et al. 2001] [Li 2001] [Köller et al. 2001] [Riemer et al. 2001] [L'Abbate et al. 2001] [Hirsch et al. 2001] [Vesterager et al. 2001] [Resgue et al. 2001] [Cunningham et al. 2001]. [Ollus 2001] provides the goal of classifications: to *help* with something.
[Turk et al. 2001] stress that for the building industry only national classification systems exist.
[Falcao et al. 2001] mention UNSPSC as *the* effort to build a comprehensive classification system for everything that can be bought or sold. (Note, UNSPSC is also sometimes referred to as an ontology).
[Cunningham et al. 2001] and [Abecker et al. 2001] try to classify information by mining the textual data and extracting enough information to classify the documents automatically.

Classification as a user-friendly view
"The information can be viewed through the optics of several classification systems" [Turk et al. 2001]. This shows the use of existing classification system as a filter that's placed over information to present it in a structure well-known to the user.

Conclusions
Classifications are made to help; that is, to help the human or the program to structure or to find information. A classification is a ready-made or evolving structure, much like a collection of "labelled boxes" in which to place information.
An alternative view is to look at classification as the process of putting information in "boxes" without any particular interest in the "boxes" itself.
Using more specific phrases like "a classification based on characteristics of ..." [Riemer et al. 2001] can improve the communication value of the term "classification".

Taxonomy

Taxonomy as a hierarchical structure
[Rutten et al. 2001] see a taxonomy as a hierarchical structure to aid the process of classifying information. The data is mostly textual. "[A taxonomy's] main goal is to systematize a gamma of various elements in a hierarchic structure" [van Schoubroeck et al. 2001]. A third paper [Riemer et al. 2001] sees a taxonomy as a tree of choices (local/global, short-term/long-term, etc.), basically using it to classify information.
A very helpful observation is made by [Simmons 2001], who equals "taxonomy" and "a consistent and repeatable analytical framework". He mentions viewpoints from which to perform a *taxonomic examination*, thereby hinting at a hierarchical structure that branches out at various levels to provide clear guidance on in which branch to classify information.

Taxonomy equals directory

[Khan 2001] equals UDDI with a taxonomy. UDDI is an internet-based directory[3] of businesses ("yellow pages"). A yellow-pages directory implies a small hierarchy (eating and drinking > restaurants > pancake restaurants). [Khan 2001] specifically mentions that his taxonomy incorporates a hierarchy: "a web directory is a hierarchical taxonomy that classifies the information".
[Dogac et al. 2001] also supports directories that are taxonomies. The first level branches out in various markets, then in businesses. The last two levels are documents (like catalogues) and data elements (the actual items).

---

[3]A directory is also a technical term for a read-optimised database. Here it is used in the sense of yellow pages or a telephone directory.

Conclusions

One thing is very clear from every citation: a taxonomy is a hierarchical structure to classify information.

A more focussed use of a taxonomy might be suggested by the use as a consistent and repeatable analytical framework [Simmons 2001]. "Consistent" can best be assured by translating it as "consistent according to data internal to the taxonomy", which fits in well with the original definition in section `First definitions'.

Ontology

Ontology just being used

Many authors at this conference just *use* an ontology, which means that they point towards items in the ontology, using it as a commonly available set of definitions or concepts [Sousa et al. 2001]. For instance to ensure that the terminology used has a common semantics.

[Dogac et al. 2001] stress that a common ontology is needed to achieve an industry-wide interoperability and [Bourdeau et al. 2001] see a need for "a high-level ontology of the construction domain to serve as a basis for knowledge indexing and retrieval".

Ontology being created as central definition set

[Fersoe et al. 2001] describe the MKBEEM project (IST) that created two ontologies: one domain ontology and an e-commerce ontology. These two ontologies together describe the concepts used.

The authors at this conference that create their own ontology as a core component of their system are more specific in what they mean by an ontology and try to define it. For example, [Abecker et al. 2001] sees an ontology as a set of descriptions of concepts. These concepts are used as definitions that are referenced. They use three ontologies.

"An ontology is an agreed (and formal) description of shared concepts in some domain which has the objective of enabling shared understanding and communication. (...) An ontology acts as a standardised reference model to support information integration and knowledge sharing" [Zwegers et al. 2001] .

"An ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents (some or all human, other artificial). (...) A commitment to a common ontology is a guarantee of understanding and consistency, but not of completeness, with respect to queries and assertions using the vocabulary defined in the ontology" [Jansweijer et al. 2001] .

Conclusions

An ontology's goal is to provide a common, referencable set of concepts for use in communication. Those concepts can be described or defined.

It is quite common to use multiple ontologies, each providing concepts from a different domain, to obtain a large enough set of concepts for meaningful communication.

NEW DEFINITIONS

Here are the definitions I will use.

Classification

"Simple classification". A grouping of entities according to some external criteria. The grouping will be quite natural, as it is mostly made from a specific viewpoint. Classification is basically a set of boxes (with labels) to sort things into. It can be used as a user-friendly view on/in a taxonomy or ontology.

Taxonomy

"Classification taxonomy" or "simple ontology". A hierarchical grouping of entities according to data internal to the taxonomy. When used as a simple ontology, the taxonomy's hierarchy should be based upon a subclass hierarchy.

Ontology

A set of well-defined concepts describing a specific domain. The concepts are defined using a subclass hierarchy, by assigning and defining properties and by defining relationships between the concepts et cetera.

When using the term "ontology" an indication should be given to the kind of ontology. A very simple ontology could perhaps better be named "taxonomy", but a heavyweight ontology should specify and advertise its capabilities lest it be grouped with the apparent majority of very lightweight ontologies.

An ontology's goal is to provide a common, referencable set of concepts for use in communication. It is quite common to use multiple ontologies, each providing concepts for a particular domain, together forming a rich vocabulary for communication.

CONCLUSIONS AND RECOMMENDATIONS

"Ontology" is nowadays a fashionable term, indicator of the fact that semantic web technologies are gaining inroads in the building and construction research.

The fashionable term by itself tends to confuse more than it specifies, because there is a large spectrum of functionality that gets grouped under the one term "ontology". Care should be taken to advertise the level of functionality of the ontology, for it is very easy to either expect too much or too little of the thing labelled "ontology".

When you label something "classification" the expectations normally match what is offered. For "taxonomies" again some care needs to be taken, at least to specify whether it is more a lightweight ontology or more a structured classification.

In the near future, more heavyweight ontologies will become more and more important with the emergence of the semantic web. The trend observed at a recent conference to call even the most lightweight classification an "ontology" makes that term less valuable, which should be avoided.

Lastly, I'd like to say a word of thanks to the reviewers who provided very valuable input.

BIBLIOGRAPHY

Andreas Abecker and DECOR consortium.
Decor - delivery of context-sensitive organisational knowledge.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Dietmar Bönke, Eckhard Ammann, and Jörg Zabel.
Knowledge engineering in virtual organisations.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Marc Bourdeau, François Giraud-Carrier, Yacine Rezgui, and Alain Zarli.
Knowledge management in the construction industry: the e-cognos approach.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Joao Falcao e Cunha, António Amador, Henriqueta Nóvoa, Ana Correia, Joao Carvalho, António Lima, and António Conde.
Internet procurement for products and services in the construction and engineering industry - the need for European standardisation.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Paul M. Cunningham and John Carolan.
The development of factwrangler- providing classification and mark-up capabilities to enhance digital content.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Asuman Dogac and Ibrahim Cingil.
Poem: a platform for an open electronic marketplace based on eco framework.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Hanne Fersoe and Andrew Joscelyne.
New opportunities for language technology in multilingual e-markets.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Tom R. Gruber.
A translation approach to portable ontologies.
In *Knowledge acquisition*, 5(2):199-220, 1993.

James Hendler and Deborah L. McGuinness.
The darpa agent mark-up language.
In *IEEE intelligent systems trends and controversies*, November/December 2000.
Abstract available on-line at http://www.ksl.stanford.edu/people/dlm/papers/ieee-daml01-abstract.html.

Bernd E. Hirsch, Jens Schumacher, Jens Eschenbächer, Kim Jansson, Martin Ollus, and Iris Karvonen.
Extended products: observatory of current research and development trends.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Bernhard Holtkamp and Rüdiger Gartman.
Matching e-commerce platform services with business models.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Wouter Jansweijer, Joost Breuker, Jan van Lieshout, Erica van de Stadt, Rinke Hoekstra, and Alexander Boer.
Workflow directed knowledge management.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Omar Khan.
The challenge of leverage on e-interactive tools for customer and supplier collaboration.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

J. Köller, B. Braunschweig, K. Irons, M. Jarke, and M. Pons.
The cape-open laboraties network: standards for interoperable process engineering software components.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Marcello L'Abbate and Ulrich Thiel.
Intelligent product information search in e-commerce: retrieval strategies for virtual shop assistants.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Man-sze Li.
Interoperability and business models for e-commerce.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Deborah L. McGuinness.
Ontologies come of age.
In Dieter Fensel, Jim Hendler, Henry Lieberman, and Wolfgang Wahlster, editors, *Spinning the semantic web: bringing the world wide web to its full potential*. MIT press, 2002.
Available on-line at http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age.html.

Martin Ollus.
Information management for networked products support.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Olaf Peters, Jörg Zabel, and Frithjof Weber.
Concept and model for an internet broker service for bidding and procurement in the tile industry.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Reinout van Rees, Frits Tolman, and Reza Beheshti.
How bcxml handles construction semantics.

In *Conference proceedings - Distributed knowledge in building*. CIB w78, 2002.
Available on-line at conference website.

Yacine Resgue, Marc Bourdeau, Abdul Samad Kazi, and Alain Zarli.
An open specification and framework for the construction dynamic virtual organisations: the osmos project.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Kai Riemer, Stefan Klein, and Dorian Selz.
Classification of dynamic organisational forms and co-ordination roles.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Huub Rutten and Steve Rogers.
Lore: language-based expertise mining.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Caroline van Schoubroeck and Herman Cousy.
Virtual enterprise legal issue taxonomy.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Stephen Simmons.
The case for immaterialisation.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Jorge P. Sousa, António L. Soares, César Toscano, and Américo L. Azevedo.
Architectures and methods for information systems promoting co-operation and innovative business processes in enterprise networks.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Frits Tolman, Michel Böhms, Celson Lima, Reinout van Rees, Joost Fleuren, and Jeff Stephens.
Econstruct: expectations, solutions and results.
ITcon, Special issue on European projects, 2002.
Available on-line at http://itcon.org.

Ziga Turk, Robert Amor, Dave Bloomfield, and Tomo Cerovsek.
Information services to enable European construction enterprises: the i-seec project.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Johan Vesterager, Peter Bernus, Jens Dahl Pedersen, and Martin Tolle.
The what and why of a virtual enterprise reference architecture.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.

Arian Zwegers, Matti Hannus, Martin Tolle, Jeroen Gijsen, and Roel van den Berg.
An architectural framework for virtual enterprise engineering.
In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce*. IOS Press, 2001.