Edinburgh Research Explorer

On the Evolution of Classifications

On the Evolution of Classifications *

Alan Bundy

School of Informatics, University of Edinburgh
Edinburgh, Scotland, UK
A.Bundy@ed.ac.uk

Abstract

The *evolution* of ontologies has been identified by W3C as a significant issue. *Automatic* ontology evolution is increasingly important to allow run-time communication between vast collections of online agents. For this application, manual, offline evolution is both too slow and too expensive.

Very many ontologies are classifications of a set of objects. In order to gain insight into how the evolution of classifications might be automated, we discuss some historic examples of such evolution. From this discussion, we extract some general principles that might be employed to help automate the ontology evolution process.

1 Introduction

The 2004 W3C Recommendation on the "OWL Web Ontology Language: Use Cases and Requirements" justifies *ontology evolution* as follows:

> *Since the Web is constantly growing and changing, we must expect ontologies to change as well. Ontologies may need to change because there were errors in prior versions, because a new way of modeling the domain is preferred, or because new terminology has been created (e.g., as the result of the invention of new technology). A web ontology language must be able to accommodate ontology revision.* [http://www.w3.org/TR/webont-req/#goal-evolution]

Many ontologies consist of classifications of a set of objects. A typical such classification will be a tree structure in which the nodes are subsets of the object set and arcs between them represent set inclusion. Given the importance assigned to ontology evolution by W3C and the ubiquity of classificatory ontologies, we thought it would be useful to investigate how some historic examples of how classifications have evolved. We hope that his analysis will reveal some general principles that might then be employed in the automation of classification evolution.

We take the stance that there is no perfect or final classification of a set of objects: *representation is a fluent* [Bundy and McNeill, 2006]. Classifications evolve with both our knowledge of the objects and the purpose for which they are employed. Neither our knowledge nor our applications are limited, so evolution must be seen as a non-terminating process. In a similar way, no definition may be regarded as final and complete, but will also evolve with our knowledge and applications. So all the definitions and classifications below must be regarded as snapshots in a process of flux.

2 The Classification of the Animals

From the earliest historical records, people have classified animals and plants, i.e., partitioned them into sets or arranged them in hierarchies. A typical example can be seen in Figure 1[1]. It consists of a tree whose nodes each correspond to a set of animals. An arc between two node represents a subset relation between the lower and the upper node.

Notice that each node of the tree in Figure 1 has some associated text. This gives some properties common to the set of animals at that node. Some example animals are sometimes also listed.

The classification of animals has evolved significantly over time. The main driver of this evolution has been to review the defining properties of each node. For instance,

- Pre-scientific classifications were based on superficial properties, i.e., mainly matters of immediate appearance or human usage, the latter being culturally dependent.

    - An example of the former is that, because whales, dolphins and porpoises lived in the sea and were fish-shaped, they were classified as fish.
    - An example of the latter is the classification of all pests under 'vermin'. An example from the classification of plants is the use of 'vegetable' to describe a selection of plant components, including some, but not all, fruits, leaves and roots, whose common property is that we can eat them[2].

---

[1] http://www.rapidonline.com/1/1/6239-classification-of-animals-wall-chart.html

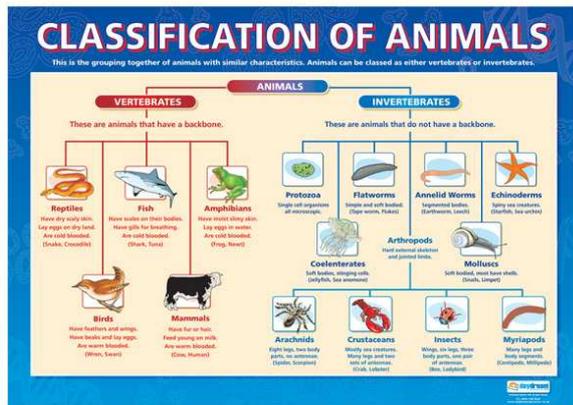[2] I'm grateful to an anonymous reviewer for suggesting this example.

Figure 1: A Classification of Animals

- In 1778, Swedish biologist, Linnaeus started to classify whales as *mammals* on the grounds that they shared significant non-superficial properties with the other mammals[3]. These properties are some of those listed as defining 'mammals' in Figure 1, that is having hair or fur, feeding young on milk, being warm-blooded, etc.

- Darwin's theory of evolution and the discovery of the fossil record led to both the inclusion of extinct animals in the classification and defining properties based on common ancestry and evolutionary descent. The hierarchical classification called *Cladistics* uses only such Darwinian properties[4].

- Most recently, the discovery of DNA has led to a further evolution in which the defining properties are based on genetic similarities. This has led to the identification of previously unrecognised distinctions. For instance, recent genetic comparison of the forest-dwelling African elephant with its savannah-dwelling cousin shows them to have diverged as much as the Asian elephant and the woolly mammoth[5]. Similarly, recent DNA studies have discovered that the closest living relatives of the beaver are scaly-tailed flying squirrels, not gophers as was previously thought[6].

- Note that this change of defining property also undermines the tree-structure of the classification. DNA dif-

---

[3] http://www.ecokids.ca/pub/eco_info/topics/whales/mammals.cfm

[4] http://en.wikipedia.org/wiki/Cladistics

[5] http://www.newscientist.com/article/mg20927942.900-ancient-genomes-suggest-africa-has-two-elephant-species.html

[6] http://www.newscientist.com/article/mg20927984.100-60-seconds.html

ferences are multi-dimensional: evolution can change any piece of DNA in multiple ways. Suppose, for instance, that gene A evolves to A1 to create one new sub-species and gene B to B1 to create another. These two sub-species may subsequently converge again in a new subspecies with *both* genes A1 and B1, making the tree into a graph. Moreover, it makes the boundaries between nodes more fuzzy. Some genetic variation has to be tolerated within a node, or each node would be singleton animal. But how much variation is necessary before a node needs to be split?

- Variant classifications may survive for a variety of reasons. For instance, simpler ones may be useful for elementary education, especially when the students cannot be expected to understand the potentially more sophisticated science informing the definitional properties of later classifications. Figure 1 is an example of such a simple classification aimed at primary school students. Some variant classifications may be aimed at specific applications, e.g., a guide to edible food, field identification for nature lovers, warnings of dangerous animals, historical interest.

From this example we can draw some conclusions.

1. We can't hope to model the evolution of classifications unless we also know the defining properties of these classifications. In the animal classification, these defining properties change from first superficial to more fundamental ones and then to the underlying DNA. Each new advance in understanding may make new defining properties available, e.g., the discovery that some animals were warm blooded and some cold blooded, the evidence of common ancestry arising from the fossil record, the discovery of DNA.

2. We have to be able to reason *about* these properties, i.e., to argue why one is a better basis for classification than another. The classification of the animals provides some possible criteria.

   - A classification may be preferred because it provides better explanatory power, e.g., DNA can potentially explain the similarities and differences between different animals.

   - A classification may be rejected because its defining properties may come to be seen as environmentally contingent rather than fundamental, e.g., fish shape is an efficient shape for animals swimming in water, which is why different kinds of animal (e.g., whales and sharks) sometimes have similar shape. Contingency may also arise because a property is defeasible, e.g., some birds can't fly, an insect may lose one of its six legs, etc. Defeasible properties do not provide reliable definitions, e.g., penguins are still considered birds even though they are flightless.

     Note that what may be viewed as fundamental and final at one stage of classification evolution may be viewed as merely contingent at a later stage. For instance, the convergent evolution of the shape of

sharks and whales, and the efficiency of fish shape are both relatively recent discoveries. So we should not be seduced into believing that the current defining properties represent a *completion* of the evolutionary process. For instance, an animal's DNA can now change during its lifetime due to genetic engineering, so even this property is defeasible.

- A classification may be preferred because it leads to successful prediction, e.g., animals found to have some of the properties of mammals may be predicted to have others, even before these other properties are observed. Perhaps live birth and hair in whales are examples. Novel but viable bacterial cells have even been created from artificially constructed DNA, after analysis predicted what the minimal viable DNA was[7].

- A classification may be preferred because it provides a finer granularity of distinction, e.g., the Linnean hierarchy contains kingdoms, classes, orders, genera and species, with wider branching and finer distinctions being frequently added, as new scientific discoveries are made. DNA provides an exponentially bigger space of possibilities yet.

- A new object may be discovered that does not fit into the existing classification, forcing a change to the classification so that it does fit, e.g., the duck-billed platypus has features of both mammals and birds. Its discovery led to the creation of a new order of mammals: the monotremes.

- Classifications may be tailored to particular purposes or usages, e.g., simplified versions for elementary education, field observation, edible *vs* inedible food, etc.

3. We have to be prepared for the shape of the classification to change if the properties do. For instance, Linnean trees may not be the best way to summarise DNA differences.

3 The Classification of the Elements

Similarly to the animals and plants, classifications of the elements go back to ancient history. The earliest classification was into four elements: earth, water, air and fire, and was due to the Greek philosopher Empedocles of Sicily[8]. The basis of this classification seems to be superficial physical appearance, although there also seems to have been a religious influence, with each element being associated with a god: Zeus, Hera, Nestis and Aidoneus.

Our modern classification originated with Mendeleev's periodic table[9] (see Figure 2[10]). Firstly, the number of classified elements changed from 4 to 92, with the original 4 elements now reclassified as mixtures, compounds or both, of the new 92 elements, or, in the case of fire, as being either a chemical reaction or photons, according to how you view it. Secondly, the new basis of Mendeleev's classification was chemical and physical properties of the elements. For instance, the elements line up in order of atomic weight: left/right, top/bottom. Each column contains elements of similar properties, e.g., the noble gases that line up in the rightmost column share the property of being chemically inert, while the alkali metals line up in the left-hand column.

Figure 2: Mendeleev's Periodic Table

Our modern understanding of atomic structure has given an explanation of why elements in the same column have similar properties. Electrons are arranged in nested shells around the nucleus[11]. There is a limit to how many electrons can occupy each shell, but this limit increases quadratically. The number of electrons occupying the outermost shell determines the chemical bonds that the element can make. So, for instance, when the last shell is full, the element is chemically inert: a noble gas. The fact that the size of each shell increases, explains why the rows of the periodic table have gaps at the top and why two separate rows are required at the bottom: the lanthanides and the actinides.

A better way to represent this would be as concentric rings of increasing size, to reflect the growing size of the shells. There have been various attempts to produce such a circular periodic table[12], but the quadratic growth of shell size makes this difficult to do while retaining visibility in the outer rings. The best attempt I could find in a short search is in Figure 3[13], but even here the lanthanides and actinides are squeezed into one entry, rather than using a bigger ring. Note that, once again, the defining properties of the classification have changed. Each element is placed in the ring that corresponds to the number of occupied shells in its electron cloud, whereas

---

[7] http://www.newscientist.com/article/
dn18942-immaculate-creation-birth-of-the
-first-synthetic-cell.html

[8] http://en.wikipedia.org/wiki/Empedocles

[9] http://en.wikipedia.org/wiki/Mendelev

[10] http://www.tutorvista.com/content/science/
science-ii/periodic-classification-elements/
modern-periodic.php

[11] http://en.wikipedia.org/wiki/Electron_
shells#Shells

[12] http://en.wikipedia.org/wiki/Alternative_
periodic_tables

[13] http://en.wikipedia.org/wiki/File:
TROPE_777x777_AB_082009.png

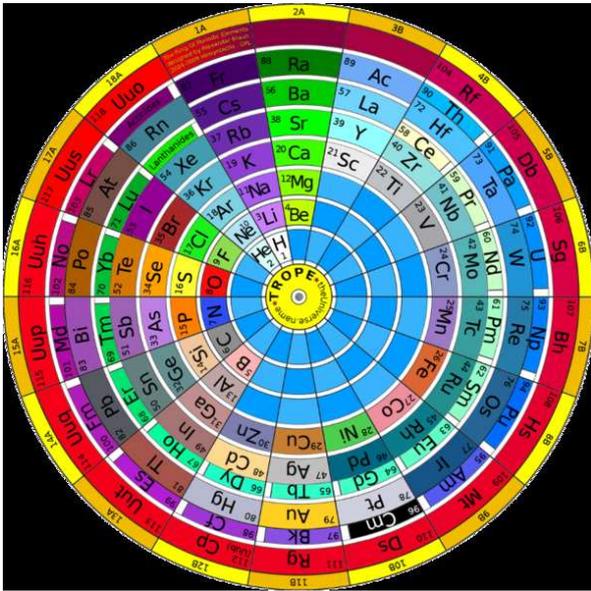in Figure 2 an element's row was determined by its chemical properties[14].

Figure 3: The Ring Of Periodic Elements (TROPE) (August 2009)

We can draw similar conclusions from the classification of the elements that we drew from the classification of the animals.

1. The change in defining properties of the elements, e.g., from physical appearance to chemical properties to atomic structure, drove the evolution of their classification. Scientific advances provided the candidate defining properties for the new classifications, e.g., separation between compounds and elements, discovery of their chemical properties and atomic weight, the quantum structure of the atom, etc.

2. To model this we'd have to reason why atomic structure is a better basis for a classification than chemical properties, which was a better basis than superficial appearance. The reasons echo some of those at work in the evolution of the classification of animals.

   - Better explanatory power, i.e., atomic structure explains chemical properties which, in turn, explains physical appearance.
   - Defining properties seen as contingent, e.g., whether an object appears at room temperature as solid (earth), liquid (water) or gas (air), depends only on its freezing and boiling points.
   - Prediction of new elements, by both Mendeleev's table and quantum mechanics.

---

[14] All these circular representations have two rings of 8 elements and two of 18, whereas a naive understanding of electron shells would suggest that there should be only one of each. It seems that some shells cannot be completely filled until the next one is started.

- A finer granularity of distinction, e.g., different kinds of 'air' in Mendeleev's table, different isotopes of the same element in quantum mechanics.
- The existence of objects that don't fit well into the existing classification, e.g., the lanthanides and actinides.

3. The move from 4 to 92+ elements was driven by chemical analysis as was the shape of Mendeleev's table, e.g., the discovery of multiple kinds of 'air' undermined its classification as a single element and led to the identification of new gaseous elements (and compounds). The move to a circular 'table' was driven by the understanding of electron shells.

4. Mendeleev's table is retained for educational purposes. Perhaps the quantum mechanical basis for a more sophisticated story would be hard to explain to students. Also, the majority of working quantum physicists subscribe to the 'Copenhagen interpretation', which rejects intuitive, non-mathematical accounts as unobtainable, misleading and undesirable, so they might not even *seek* a better classification.

4 Classification in Action in Astronomy

It's instructive to witness classifications being formed. We're historically lucky that several such classifications are currently being hotly debated, especially in Astronomy.

The case that has received the most public attention is whether Pluto should be classified as a planet[15]. What drove the debate was the discovery of similar objects to Pluto in the outer solar system, e.g., Chiron, and the expectation that many more such objects would be discovered in the future. It's interesting that the principal arguments on either side of the debate were more 'political' than scientific.

- Those arguing for the demotion of Pluto to a new classification of *dwarf planet* were mostly motivated by wanting to keep the designation 'planet' as something special. They did not want to open the door to huge numbers of objects being called planets.
- Those arguing for the retention of Pluto as a planet were mostly driven by tradition, i.e., Pluto had been classified as a planet since 1930, and they were reluctant to see this decision revoked.

The arbitrary nature of the final decision is emphasised by the method of decision: a vote, on August 24 2006, by the International Astronomical Union. However, what they voted for was some new defining properties of a planet, which would exclude Pluto, Chiron and other anticipated planetary candidates. These were:

1. The object must be in orbit around the Sun.

2. The object must be massive enough to be a sphere by its own gravitational force. More specifically, its own gravity should pull it into a shape of hydrostatic equilibrium.

3. It must have cleared the neighbourhood around its orbit.

---

[15] http://en.wikipedia.org/wiki/Pluto

Pluto (and Chiron *et al*) fails property 3. Its mass is too small to clear out other objects in its neighbourhood. One might speculate that property 3 was added *precisely* to exclude Pluto, since it seems a bit esoteric compared with the other two properties.

A similar debate may be brewing over the separation between star clusters and galaxies[16]. Star clusters are collections of stars that formed simultaneously from the same gas cloud, whereas galaxies contain enough gas to form several generations of stars, but neither have a formal definition. Recent observations of the star cluster Omega Centauri, however, suggest that it might contain multiple generations of stars. Two astronomers have recently launched an informal poll to ask astronomers what they think the defining properties of a galaxy should be.

Again, we see voting techniques being used to settle a classification problem. This recognises that, ultimately, although scientific properties are used to define a classification, the choice of those properties is essentially a political decision. We see similar drivers as in the case of Pluto:

- A galaxy is something special. We don't want to grant its status too readily.

- We should respect tradition and not change the status of an object too readily.

In this case, however, both these arguments are pushing in the same direction: to retain the classification of Omega Centauri as a star cluster. I expect the polled astronomers to agree on an additional defining property of galaxies that excludes Omega Centauri, in the way that property 3 above excludes Pluto as a planet[17].

5   Conclusion

We have looked at a few examples of how classifications have evolved historically. This has lessons for us in the automation of classification evolution.

Firstly, classifications are defined by properties of the objects being classified. Objects with similar properties are placed into the same class. Moreover, objects with only slight differences in their properties are classified closer together than those with major differences. Advances in science may make new kinds of defining property available.

Secondly, to emulate classification automation, a system will have to reason about rival defining properties, i.e., deciding that one is better than another. Common preference criteria are:

- Better explanatory power of the new classification.

- Defining properties initially seen as fundamental are subsequently seen as contingent and, therefore, rejected in favour of more fundamental ones.

- Successful prediction from the new classification.

- Finer granularity of distinctions in the new classification.

- The discovery of objects that don't fit into the old classification.

Sometimes, however, defining properties are chosen not by these scientific criteria but on 'political' grounds, e.g., a preference for reserving some classes to a few privileged objects or a conservative attitude to change. We saw these at work in the astronomical examples of §4, but they can also be detected in other examples. For instance, the hierarchical structure of the animal classification is partly driven by a desire not to have too many or too few animals in each class. We also see the power of tradition in the continued use of Mendeleev's periodic table, even though a classification based on quantum mechanics might better fit the scientific criteria. Another example of tradition can be seen in classification of star temperatures. The letters O, B, A, F, G, K, and M classify stars from hottest (O) to coldest (M). The choice of letters arises from an earlier, incorrect assignment, in which they were in alphabetical order, ABCDEF..., where A was the hottest[18].

Sometimes, classifications are devised for specific purposes, e.g., education, particular usage, field observation in the absence of sophisticated instruments, etc. In these cases the reasons for preferring particular defining properties will also be informed by those specific purposes.

Thirdly, the types of the defining properties used will determine the structure of the classification. We have seen partitions, hierarchies, tables and concentric rings, for instance, and the evolution from one to another. [Kemp and Tenenbaum, 2008] describes some interesting work automating the form of a classification from the defining properties. Such a system might serve a useful role in classification evolution, i.e., by automatically constructing a new classification once its defining properties have been determined.

Fourthly, the evolution of a classification may also require an enrichment of the representation language. This can arise not only because of an increased complexity of the structure of the classification, e.g., from a mere naming of the elements to the periodic table. It can also arise from the need to express more complex defining properties. The defining properties used in the simple animal classification of Figure 1 are unary predicates, but Cladistics requires binary and ternary relations of descent and common ancestry. A classification based on DNA requires functions returning a complex structure of chromosomes, genes and double helices of nucleotides. Similar remarks can be made for the classification of the elements. The representation of the defining properties required in Astronomy requires complex geometric concepts, such as a sphere and an orbit and its neighbourhood. Note that an orbit, for instance, is naturally represented as a function from time to 3D position, which takes us into higher-order logic. Note that the representation language used to express the defining properties is frequently more sophisticated than that used to describe the classification itself. This suggests that ontology evolution will itself require a more sophisticated language than the one used to describe the ontology being evolved.

Fifthly, what drives the process of classification evolution? It may arise from within the classification, e.g., the failure to

---

[16]http://www.newscientist.com/article/
dn20026-when-is-a-group-of-stars-not-a-galaxy.
html

[17]This prediction has been subsequently confirmed by the outcome of the vote.

[18]I'm indebted to an anonymous reviewer for this example.

classify a newly discovered object or to be useful for a new application. More likely, however, is that it will arise from an evolution of knowledge leading to better defining properties, i.e., properties that are more explanatory, fundamental, predictive and/or finer grained.

References

[Bundy and McNeill, 2006] A. Bundy and F. McNeill. Representation as a fluent: An AI challenge for the next half century. *IEEE Intelligent Systems*, 21(3):85–87, 2006.

[Kemp and Tenenbaum, 2008] C. Kemp and J.B. Tenenbaum. The discovery of structural form. *Proceedings of the National Academy of Sciences of the United States of America*, 105(31):10687–10692, August 2008.