Jens-Erik Mai
Royal School of Library and Information Science, Copenhagen, Denmark

Likeness: A Pragmatic Approach

Abstract: The paper discuses the concept of likeness in classification theory. It shows that the concept has not been adequately defined in the literature and suggests an alternative understanding of the concept. This understanding is based on domain analysis and pragmatism.

1. Introduction

Classifications are generally defined as the organization of documents according to likeness. It is further assumed that documents are organized according their characteristics, and that those documents, grouped together in the same class, are alike in the sense that they share some specific characteristics. However, before this conclusion can be reached, a definition of what documents' characteristics are is needed. The characteristics of documents on which classifications are typically based, are the documents' subject matter. This paper goes into this discussion from the standpoint that documents do not have a subject, but are given a subject description according to their use and context. Likeness, therefore, is not a stable factor of a classification, but is dynamic in nature.

Since the early days of classification theory, likeness has been emphasized as the principle which underpins any theory of organization of documents. However, the principle has some major flaws since it builds on the idea that some documents are alike.

Documents can be organized according to likeness in various ways. Documents by the same author could be grouped together; documents that cite the same sources could be grouped together; documents with similar occurrences of words could be grouped together, etc. In such cases the organization will be (relatively) stable in nature. However, of particular interest, are cases where documents are organized according to likeness in subject matter. In such cases it must be assumed that documents grouped together in a class share subject matter. This general idea is seen many places in the literature, for instance in Lois Mai Chan's textbook on cataloging and classification where she notices that one of the objectives of library classification is, "to group all works of a kind together" (Chan, 1994, 259). She further states that this, "requires the grouping of like materials together on the basis of chosen characteristics" (Chan, 1994, 259), such that the user can, "retrieve a group of related items" (Chan, 1994, 260). The material can be grouped according to a number of chosen characteristics; however, "subject is the predominant characteristic for grouping" (Chan, 1994, 260). Another widely used textbook states the same basic assumption, although the wording is a bit different, namely, that, "the purpose of classification is to bring related items together" (Wynar, 1992, 317), such that the user, "can find related works in one location" (Wynar, 1992, 317). The idea is most clearly expressed by Richardson (1935, 1) who stated that, "the 'putting together of like things' is . . . the fullest and most exact" definition of classification, and by Buchanan (1979, 9), who stated that, "classification is the act of bringing like things together."

In order to fully understand and use the concept of likeness, we need to clarify what is meant by the concept. The major issue, therefore, is to understand on what basis the grouping is made, how the characteristics are chosen – in short, how the likeness and unlikeness of documents are determined.

2. Logical Division

The determination of which documents to bring together, and which to separate has, in the literature, often been referred to as a logical decision. The study of logic, therefore, is often regarded as the basic study for understanding classification. Sayers (1915, 16), for instance, stated that, "Classification is a department of logic, and every step in the construction of a classification scheme is referable to that science."

He went on to say that a student of classification should to be acquainted with logic, and that, "...some such book as Jevon's *Primer of Logic* should be the minimum preliminary reading; or, better still, the same author's *Elementary Lessons in Logic*". (Sayers, 1915, 17). Only after the student has learned the basics of logic will he or she be able to fully understand and appreciate bibliographic classification.

Two of the most commonly mentioned logical principles are those of exclusivity and exhaustivity. Exclusivity states that classes on the same level should be distinct, such that documents placed in one class could not also be placed in another class. This is achieved by applying, "only one characteristic of division . . . at the time" (Buchanan, 1979, 53). The other principle is that of exhaustivity, which states that, the "division by a characteristic must be exhaustive – we must include all the classes produced by each characteristic, or our scheme will be incomplete" (Buchanan, 1979, 55). This idea can be traced back to Porphyry (c.232-c.304) who is known for explicating Aristotle's categories and thereby developing what is known as the "Tree of Porphyry."

The tree of Porphyry is the principle that a given set of objects at the highest genus can be divided into mutually exclusive and collectively exhaustive subordinate genera. Each class is divided by adding differentiae, which is a set of opposites. This is continued until the lowest possible species are reached and the set of objects cannot be divided further.

These principles have worked fairly well in the classification of natural kinds, as, for instance, in Linnaeus' classification of living things. The reason is that the characteristics chosen, such as the shape of a fruit, are easy to perceive and describe. Furthermore, all biologists and botanists would agree on the interpretation of the characteristics (Lakoff, 1987). Such taxonomy does not intend to analyze the meaning of the terms, but is merely a classification of kind of things. The chosen characteristics by which the genus is divided into genera are properties of the things classified, and, as such, the characteristics are subject to inspection. However, the users of such taxonomy know that the use of the classification requires some sort of interpretation. That is why a zoologist would not dispute a statement like 'this cat has three feet,' since he or she knows that there can be handicapped cats. But he or she would still classify cats as four-footed mammals and he or she would still say that the property of being four-footed belongs to cats, but he or she would not say that cats are four-footed necessarily or analytically (Eco, 1984). In other words, nothing specific is said about individual cats in such a classification.

Eco (1999) mentions a few good examples where classifications of natural things run into problems. One problem is that an object could have characteristics such that two entirely different classifications could be justified. Ayers Rock, for instance, is a mountain in the sense that it displays all the characteristics that we commonly attribute to mountains, however Ayers Rock is a stone in the sense that it is "a monolith planted in the ground as if a giant had hurled it" (Eco, 1999, 225). In such a case there must be consensus about what characteristics should be chosen in the classification. Another problem is the case where some natural kind simply does not fit into a classification at all. Eco (1999, 58) mention the platypus as such an example:

> The platypus is a strange animal. It seems to have been conceived to foil any classification, be it scientific or popular. On the average about fifty centimeter long and roughly two kilos in weight, its flat body is covered with a dark-brown coat; it has no neck and a tail like a beaver's;

it has a duck's beak, bluish on top and pink beneath; it has no outer ears, and the four feet have five webbed toes, but with claws; it stays underwater (and eats there) enough to be considered a fish or an amphibian. The female lays eggs but "breast-feeds" her young, even though no nipples can be seen (the male's testicles cannot be seen either, as they are internal).

Any classification would most likely have to give up here in the sense that classes should be mutually exclusive. This animal possesses so many unique characteristics that any classificationist would be in trouble. Applying the principles of logical division would not help the classificationist much. He or she would have to make a choice as to how to classify the animal. The very structure of the classification is based on nothing else but the classificationist's view of the world, although it is often claimed that the order of nature is a classified order, and the classificationist merely reflects this order. To that claim, one must ask, "if it is, who classified it?" (Broadfield, 1946, 31).

Nonetheless, in most cases there will not be any problems, and it would be rather obvious to which class a given object belongs. The reason is that the classificationists and users of the systems would agree on how to interpret the structure of the system, and the rules for classifying in the system. In cases where there are disagreements between the classificationists and/or the users, they would be able to settle such a disagreement by inspecting the objects classified, and reach a consensus on what characteristics to classify a given object by. In that sense there are no differences between classifications and language, and classifications must be regarded as social constructs independent of nature and physical reality. Therefore, nothing specific is said about the individual objects that are classified in a specific class. All which is said is that they share certain chosen characteristics.

This idea has been transferred to the classification of documents. "Classification . . . is about the discovery and display of relationships," as Buchanan (1979, 14) states. When we classify natural kinds we generally assume that there is an order to be discovered and displayed. This assumption is questionable, however. But in most cases it is possible to reach an agreement on the classifications and this agreement sometimes has been used as proof that the classification displays reality. Nevertheless, the idea that classification is used to discover and display relationships underpins theories in bibliographic classification research. However, these relationships are not discovered but constructed.

It can therefore be concluded that there is one more thing to be added to Miksa's (1998) list of ideas that the library classificationists took from the older movements to classify knowledge and sciences. The notion of classifying on the basis of resemblance or likeness was brought over, and bibliographic classification theory is based on the idea that the best way to structure documents is by the use of logical principles and that these principles form the core knowledge of classification. However, a closer look at these principles reveals that much more than logical principles are in action in the construction of classifications.

3. Subjects of Documents

It is one thing to classify natural things, and another to classify documents according to their subject matter. As mentioned earlier, documents can be organized in various ways. They could collocated according to shared author, shared citations, shared occurrence of words, etc. But, of interest in this paper, is the organization of documents according to shared subject matter. This notion requires that it is possible to determine the subject matter of documents, such that the documents with the same subject matter would be collocated. Patrick Wilson (1968, 89) once noted:

> The notion of the subject of a writing is indeterminate, in the following respect: there may be cases in which it is impossible in principle to decide which of two different and equally precise

descriptions is a description of the subject of a writing, or if the writing has two subjects rather than one.

If it is impossible to give a precise description of a document's subject matter, it makes little sense to claim that documents are arranged according to likeness in subject matter. Literary theorists have discussed similar issues and have recently begun to argue that the meaning of a text is what the addressees find in it by virtue of their own systems of expectations, or as Stanley Fish (1980, 2-3) argues:

> [I]f meaning is embedded in the text, the reader's responsibilities are limited to the job of getting it out; but if meaning develops, and if it develops in a dynamic relationship with the reader's expectations, projections, conclusions, judgments, and assumptions, these activities (the things the reader *does*) are not merely instrumental, or mechanical, but essential, and the act of description must both begin and end with them.

According to Fish's understanding, a text does not have a meaning. Instead, the reader creates meaning as the text is read. A reader, therefore, does not respond to the meaning of a text. The reader's response is the meaning of the text. According to this logic, a document does not have a subject, but is given a subject by the reader. The classifier's task, therefore, is not to get the subject out of the document, but to create the subject and to express this interpretation in the classification system. This interpretation, however, should be based on the potential users' work domain.

A classification is a structured view of the world in general or a view of a specific domain of knowledge. The domain can be formed as a field of study, e.g., biology, economics, or sociology, or the domain can be formed as an otherwise shaped group of people with a specific purpose, e.g., interdisciplinary studies, a company, or an organization. A domain, therefore, is first and foremost a group of people forming a discourse community.

Wilson (1968, 20) noted that, "much . . . of the reading we do is purposive"; we therefore seek the "best textual means" (Wilson 1968, 20) for our purpose. What we seek is not just any document treating a specific subject, but some documents that, for instance, will help us solve a problem. Therefore, "there need be no relationship known to logic between a text and a problem it helps someone solve" (Wilson 1968, 49), and classifications would be much stronger if they reflected the domain's discourse and use of documents rather than attempted to discover and display relationships between documents.

4. Conclusion

The application of the concept of likeness in bibliographic classification clearly has some problems. Even if the idea that documents that belong to the same class share the same subject matter is accepted, it is not clear in what respect these documents are distinct from the documents in the class next to them. It could be argued that the documents have a central or predominant subject, and that they are classified according to this. However, the determination of the subject matter of documents is of such a fundamental interpretive, and, therefore, variable, nature that nothing specific can be said about the outcome of it (Mai, 2000). Another classifier at another time might have seen some other characteristics in the document, and might have classified the document differently. To argue that documents have a predominant subject, and are classified according to this is clearly senseless. Furthermore, any representation and interpretation of a document's subject will be based in, and influenced by, a social and cultural context.

To fully understand how documents are collocated in a classification system, and how the classification system was structured in the first place, one needs to understand the difference (or lack of difference) between what Wilson (1968) called descriptive power, and

exploitative power, and what Eco (1984; 1999) called dictionary competence, and encyclopedic competence. In the first sense of classification, we limit our task of classifying simply to registering that a given document belongs to a certain node in the classification. The criteria of likeness of documents that belong to a certain class, can be established objectively, and neutrally and it requires no special skill to assign documents to the given class. In the second sense we employ knowledge about the classes, and knowledge about the content and use of the documents in assigning the documents to a given class. The drawback is that the criteria of likeness are much less obvious and it requires a special kind of knowledge to classify documents. It requires that the classificationists and the classifiers have in-depth knowledge about the domain and language used in the domain. In this sense, classification is an art that goes beyond any formal rules of logic and science.

The characteristics used in the classification and collocation of documents include interpretations of the documents' subject matter. Likeness, therefore, is a concept that is closely related to the pragmatics of the language used in a given domain. The understanding and mastery of the domain and its language is the gateway to any work in classification and representation of knowledge.

References
Broadfield, A. (1946). *The philosophy of classification*. London: Grafton & Co.
Buchanan, B. (1979). *Theory of library classification*. London: Clive Bingley.
Chan, L.M. (1994). *Cataloging and classification: an introduction*. New York: McGraw-Hill.
Eco, U. (1999). *Kant and the playtapus*. London: Secker & Warburg.
Eco, U. (1984). *Semiotics and the philosophy of language*. London: MacMillian.
Fish, S. (1980). *Is there a text in this class?* Cambridge: Harvard University Press.
Hjørland, B. (1997). *Information seeking and subject representation*. Westport: Greenwood Press.
Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.
Mai, J.-E. (2000). *The subject indexing process*. Ph.D. diss., University of Texas at Austin.
Miksa, F.L. (1998). *The DDC, the universe of knowledge, and the post-modern library*. Albany, NY: Forest Press.
Richardson, E.C. (1935). *Classification*. New York: H.W. Wilson.
Sayers, W.C.B. (1915). *Canons of Classification*. London: Grafton & Co.
Wilson, P. (1968). *Two kinds of power*. Berkeley: University of California Press.
Wynar, B.S. (1992). *Introduction to Cataloging and Classification*. Englewood, CO: Libraries Unlimited.