

The Concept of Subject in a Semiotic Light

Jens-Erik Mai

Royal School of Librarianship, Copenhagen, Denmark

Abstract

The paper argues that if digital objects are considered signs, the process of representing the subject matter of digital objects could be analyzed in a semiotic framework. The paper discusses and introduces such a framework.

The framework is based on two elements of the American philosopher Charles Sanders Peirce's semiotics, viz. the notion of 'unlimited semiosis' and the categorization of signs. Peirce has discussed how meaning is derived from signs and how new meaning is produced on the basis of earlier interpretations. He has also defined and categorized ten types of signs. It is in this categorization made clear that signs differ in their level of interpretation required to interpret them, and in how they relate to the objects which they represent.

The framework relates Peirce's notion of unlimited semiosis to the subject indexing process and categorizes each element of the process in Peirce's categorization of signs. It is concluded that more research on the subject indexing process is critically needed. Especially research on the first step of the process in which the meaning of the digital object is determined.

INTRODUCTION

One of the key functions of library and information services is to provide access to information to based on users' requests for knowledge. Knowledge can be stored in a wide range of information bearing objects such as text, image, sound and multimedia, and as technology develops more people gain access to the objects, through different media. We will here analyze the processes and problems associated with determining the subject matter of an information bearing object.

In 1968 Patrick Wilson noted that "the world is full of writings". "In libraries, archives, offices, and attic trunks is an enormous and rapidly increasing mass of writing material of all sort" he continued (Wilson 1968, p. 1). This mass of writings has increased further since. And with a still higher speed. Especially as we see the development of new media and new ways of communicating. The Internet is an easy example here. The amount of databases available is growing each day, and so is the amount of records contained in these databases. It is popular to note that we are moving into the information age, which generally means that more information is available 'at our fingertips'. One of the challenges in this new environment is to represent the subject content of digital objects.

We will here analyze key problems associated with representation of the intellectual content of digital objects. It should be noted that the analysis is based on the premise that the representation is conducted by a human indexer.

The analysis will be build up around an understanding of the process of representing the intellectual content of documents, which has been suggested by Miksa (1983). This framework is general enough to include any digital object – text, image, sound and multimedia. The analysis will use the American philosopher Charles Sanders Peirce's *semiotics*, especially his notion of 'unlimited semiosis' and his ten categories of signs.

A few other writers have related the study of semiotics to the information science field, although all quite different from our approach. Blair (1990) argued that a foundation of document representation must be based on theories about meaning and language. He found that semiotics to a certain degree could be useful in information retrieval studies. Brier (1996) found that semiotics together with second-order cybernetics and Wittgenstein's pragmatic philosophy of language could make up the foundation of the field. Gluck (1996) attempted to make a link between the sense-making concept and Peirce's triadic sign concept. Warner (1990) noticed that there is a uninvestigated

relation between information science and semiotics. These have used semiotics to discuss the fundamental philosophical and theoretical foundation of the field. We here attempt to use semiotics to analyze a practical problem.

ANALYSIS OF DIGITAL OBJECTS

The process of representing the subject matter of information bearing objects has been reported to take either two steps (Frohmann 1992; Petersen 1994), three steps (Farrow 1991; Miksa 1983; Taylor 1994) or four steps (Chu and O'Brien 1993; Langridge 1989).

The two step procedure consists of one step in which the subject matter is determined, and a second step in which the subject is translated and expressed in the indexing language. In the three step procedure there is a step between these two steps in which the subject matter is explicit or implicit formulated. In the four step procedure the translation of the subject matter into the vocabulary of the indexing language consists of two steps. The indexer first translates the subject matter from his/her vocabulary into the vocabulary used in the indexing language. Then the indexer constructs the subject entry in the indexing language in form of index terms, a class mark or a subject heading.

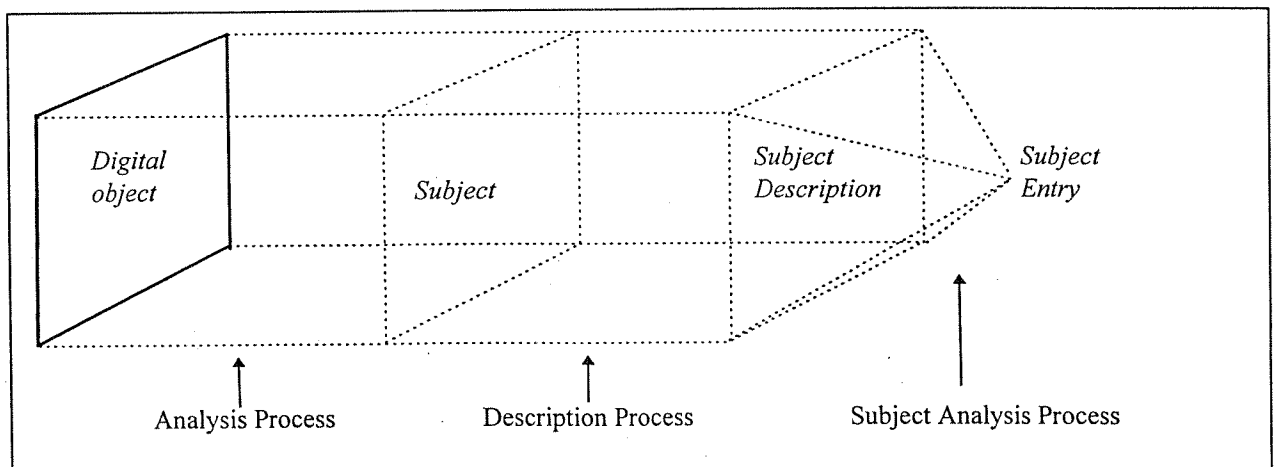
For experienced indexers and catalogers these steps may take place almost simultaneously. It may however be useful to operate with more steps when analyzing the process.

We will here use the three step model. We do not find it neither useful nor possible to split the translation of the subject matter into two separate steps.

The subject indexing process consists of three steps and four elements. The first element is the *digital object* which is under examination. Element two is the *subject* of the digital object, this element could be present only in the mind of the indexer. Element three is a *formal description of the subject*, which could be translated into writing. The fourth, and last element, is *the subject entry* which has been constructed in the indexing language.

The three aforementioned *steps* link the four elements. Step one is the analysis of the digital object for its subject, the *Analysis Process*. The second step is the formulation of an indexing phrase or subject description, the *Description Process*. The third step is the translation of the subject description into an indexing language, the *Subject Analysis Process*.

Figure 1. The subject indexing process



Miksa (1983) has discussed what a subject heading's referent is. He argued that the literature tends to answer this question by stating that "a subject heading should express or match in some essential way the topical content of a work" (Miksa, 1983, p. 5). He argued that this correspondence between a subject heading and the subject of document only has casual correspondence, since the subject matter of documents can not precisely be measured. Miksa explains this by introducing a model of the subject indexing process (figure 1, modification of Miksa, 1983, p. 6). It is in the model made clear that subsequent elements of the process, are based on preceding elements. Miksa hereby implies that the referents of each element in the process are based on interpretations of former elements, viz. that the referents of the different elements of the process are highly individual interpretations.

Miksa argued that the task of a subject cataloger is to move from the document to the subject entry, such that the subject cataloger devises a subject name for the subject content of the document. In other words, the task of the subject cataloger is to transform the content of the digital object to a representation of the digital object. This process is successfully when the subject cataloger has determined a "name or names . . . (which) suggest, represent, fit, match, etc., . . . the supposed substantiality of the topical content" (Miksa, 1983, p. 7).

How to find the subject

Manuals on classification and indexing are generally uninformative about how one identifies the subject of a document. They merely recommend examination of titles, tables of contents, scanning of chapters' headings, examination of forewords and introductions (Taylor, 1994).

It is almost impossible to formulate guidelines on how to determine the subject of given digital object. Bates (1986, p. 360) has observed that "it is practically impossible to instruct indexers or catalogers [on] how to find subjects when they examine documents. Indeed, we cataloging instructors usually deal with this essential feature of the skill being taught by saying such vague and inadequate things as 'Look for the main topic of the document'." The same problem is in effect for other media as well.

SEMIOTICS

Two traditions of the study of signs can be identified: a European and an American. The European tradition is based on the work of the French linguist Ferdinand de Saussure (1857-1913), this school is usually named *semiology*. The American tradition is based on the work of the American philosopher Charles Sanders Peirce (1839-1914), and is called *semiotics*. Both traditions are concerned with the meaning of signs, but where Saussure was concerned with semantics, Peirce was concerned with how signs are attributed meaning. Although there have been attempts to define a unified theory of semiotics (ex. Eco, 1990) the two traditions are distinct, even though they might complement each other. Saussure's theory is a theory on how to derive meaning from words, Peirce's theory, on the other hand, is about how signs are attributed meaning. Dines Johannsen (1985, pp. 225-6) explains:

As a contra distinction to the concept of sign of continental structuralism (Saussure, Hjelmslev), defining the sign as an immanent solidarity between two formal entities (an element of expression and one of content). Peirce conceives the sign as an element in a signifying process.

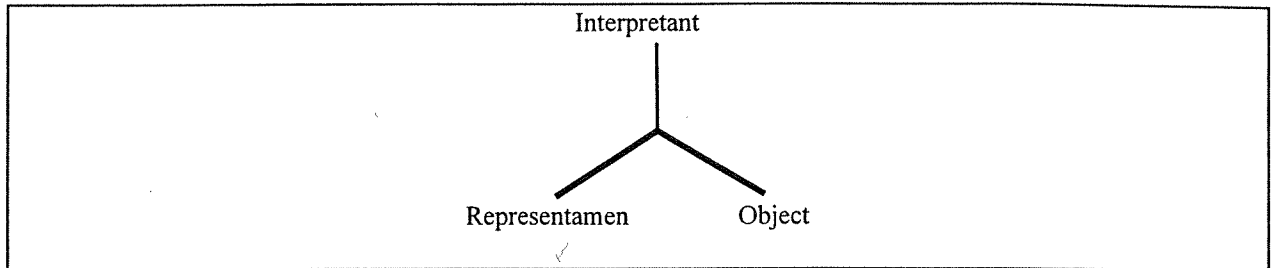
Hence, Saussure operated with a two-sided concept of the sign. He suggested that words are not merely names that represent *things*, but are expressions that *stand for* some content. By this he separated words and their content. Saussure argued against the notion that the content of words is an inherent quality in them, as it was the norm in earlier linguistics. Saussure argued that the connection between a word and its content is *arbitrary*. His theory is centered around how to *derive* meaning from words.

Peirce defined a sign as a relation among three entities, the sign itself, the referent of the sign, and the meaning which is derived from the sign. He operated with a three-sided, or a *triadic*, concept of sign, which he defined as (Peirce, 1955, p. 99):

A sign, or *representamen*, is something that stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign that it creates I call the *interpretant* of the first sign. The sign stands for something, its *object*. It stands for that object, not in all respects, but in reference to a sort of idea.

Peirce's triadic concept of the sign is best represented in this Y-leg model:

Figure 2. The Y-leg model



The bold line from *representamen* to *object* indicates the connection between the primary sign (the representamen) and its referent (the object). The connection between these two entities is the meaning of the representamen. Representamen is the primary sign [1], the physical entity, which we see or interpret. The object is the range of ideas which the primary sign refers to. The interpretant is the new sign which is created by the person who sees the sign. A distinction between the three units in the sign is thereby stressed.

The connection between the representamen and its object can only be made by the *interpretant* – which is a new and more developed sign. The interpretant is *not* a person who interprets the sign, but is the sign that is produced from the representamen. The connection between the object and the representamen is made by a person, and the interpretant is created in the mind of that person.

Since we are here interested in theories on how signs are attributed meaning, Peirce's theory is of interest, although it may be supplemented with Saussure's.

This model of how a new sign is created from another sign, can be useful in studying the subject indexing process. The digital object is the representamen, since the digital object is the item for analysis, the primary sign. The digital object contains some ideas, which it refers to, these ideas are the 'object' of the digital object. The analysis of the digital object, or the result of the analysis process (or, the first step in the subject indexing process), is the interpretant. Thereby a distinction between the *digital object* itself, its *ideas* and the *subject* of the digital object is made.

It is in a semiotic understanding assumed that there is a distinction between the digital object, the range of ideas contained in the digital object, and the subject of the digital object. There is therefore not any predicable relationship between the subject entry and the ideas which the digital object represents.

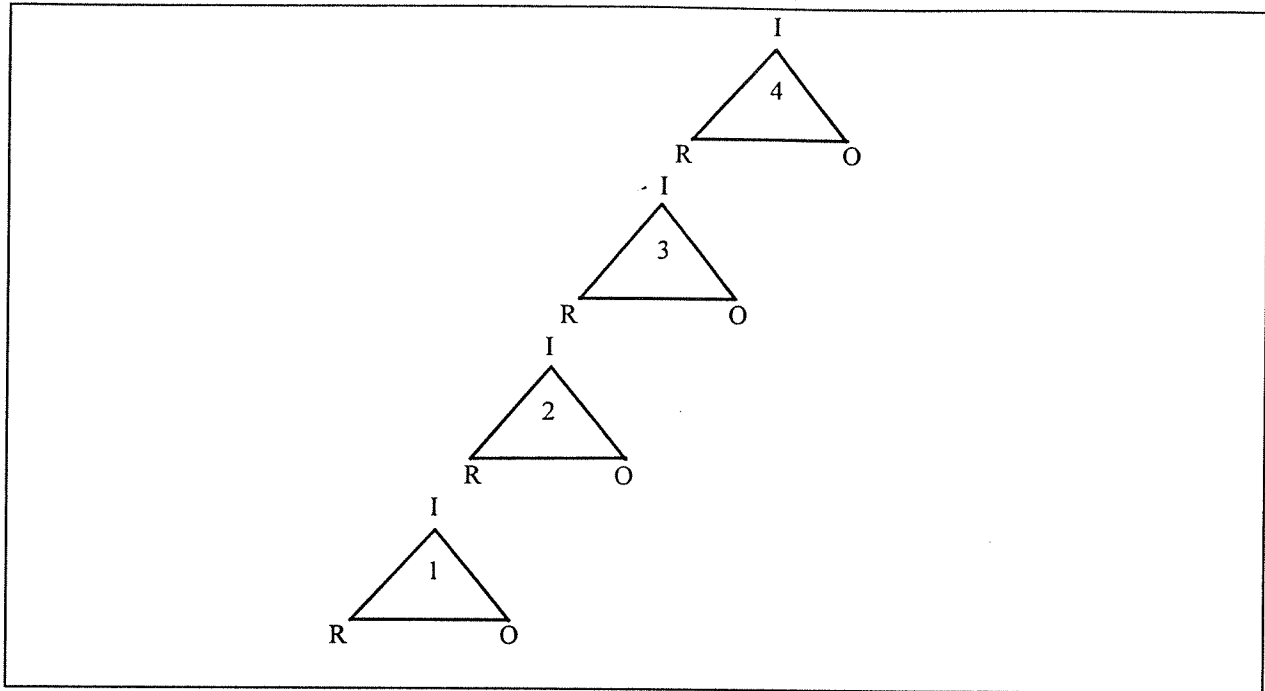
Unlimited Semiosis

As a new sign develops from another sign, 'semiosis' occurs. This process can continue infinitely, a new, or more developed, sign will always be produced when a primary sign is interpreted. Peirce called this process 'unlimited semiosis'.

When the concept of 'unlimited semiosis' is applied to the subject indexing process, the analysis of the digital object (the first step), develops a new sign, this new sign is the digital object's subject. The understanding of the digital object's subject develops yet another sign, which is the subject description. The subject description produces another sign, the subject entry.

When each *element* in the subject indexing process is regarded as a sign, each *step* is a link in the unlimited semiosis process. Unlimited semiosis can graphically be represented as:

Figure 3. Unlimited Semiosis



The triangles in the figure will continue to generate new triangles, there have been triangles before these and there will be triangles after these. An understanding of something is based on an understanding of something else, and will always generate another understanding.

When we apply the model of unlimited semiosis to the subject indexing process, the three steps in the subject indexing process will change as follows:

- Triangle number one represents the first steps of the subject indexing process, the *analysis process*. Representamen (R) is the digital object, object (O) is the content or idea of the digital object, and interpretant (I) is the analysis of the digital object.
- Triangle number two is the *description process*, the analysis of the digital object which in triangle one was interpretant (I), has now changed to become representamen (R). Representamen is therefore the subject of the digital object, object (O) is to the idea of the subject, and interpretant (I) is the subject description.
- Triangle number three is the *subject analysis process*. The subject description changes from interpretant (I) in triangle two to representamen (R) in triangle three. Object (O) is the idea represented in the subject description, and interpretant (I) is the subject entry chosen to represent the digital object.

This analysis stresses that the *object* is different in the three steps. What the digital object refers to is not necessarily the same as what the subject, the subject description or the subject heading refers to. This could to a certain extend theoretically explain the uncertainty which is connected with the subject indexing process.

The semiosis process will continue after the subject indexing process has been completed. The user or intermediary will create an understanding of the subject entry, which will have a certain object and create a certain interpretant. The user's information need represent another sign which will generate new signs as the user interacts with the information retrieval system.

Sign Trichotomies

Peirce attempted to categorize signs into a number of groups. A popular categorization attributed to Peirce is *icon*, *index* and *symbol*. This categorization of signs is based on how a sign points out what the sign is about. The icon on the basis of resemblance (like the sign on a bathroom door), an index points to what the sign refers to (like smoke to a fire) and, a symbol is based on a convention (like language).

This is a simple representation of Peirce's categorization, who enumerated ten categories of signs. He did this by defining three modes of each aspect (interpretant, representamen and object) of the sign. These are then combined to ten categories.

The three modes of each aspect of the sign are based on Peirce's phenomenology, in which he argued an understanding of the world, which divides the world into three modes of phenomena.

Peirce states that there are three modes of being: the being of positive qualitative possibility, the being of actual fact, and the being of law (or conventions). Peirce named these, firstness, secondness and thirdness. Generally they can be said to be:

- Firstness: *Monadic* (properties).
- Secondness: *Dyadic* (relations between two phenomena).
- Thirdness: *Triadic* (mediated relations between three phenomena).

Firstness is the mode of being that consists of the category of qualities of phenomena, such as red, bitter and hard. These are monadic properties. Members in this category of modes of being are regardless of anything else, they are seen as qualitative possibilities. The understanding of these qualities is reasonable alike in all humans. That is, there is general agreement on what is red, bitter or hard (Peirce, 1955, pp. 80-81):

Among phanerons [2] there are certain qualities of feeling, such as the colour of magenta, the odor of attar, the sound of a railroad whistle, the taste of quinine, the quality of the emotion of contemplating a fine mathematical solution, the quality of feeling of love, etc. I do not mean the sense of actually experiencing these feelings, whether primarily or in any memory or imagination. That is something that involves these qualities as an element of it. But I mean the qualities themselves which, in themselves, are mere maybes, not necessarily realized.

Secondness is the mode of being that consists of how other objects are. This mode of being is represented by dyadic properties, the actual facts about the phenomena. These properties make it possible to recognize and identify the phenomena directly. Secondness can be illustrated by the situation in which a person leans against a locked door, the person concludes from the impossibility of opening the door that somebody locked it (Peirce 1955, p. 88):

Some writers insist that all experience consist in sense-perception; and I think it is possibly true that every element of experience is in the first instance applied to an external object . . . We perceive objects brought before us; but that which we especially experience -- the kind of thing which the word "experience" is more particularly applied -- is an event. We cannot accurately be said to perceive events.

Thirdness is the triadic relation between something first and something second, which tells us something third. What we get from signs is *meaning*. Meaning is not inherent in signs, but something we *make* from signs. Language is based on habits, or more correct; habits become such that people think of them as laws. Thirdness is the class of these habits, which is the vagueness in language (Peirce 1955, 91):

It is impossible to resolve everything in our thought into those two elements [of firstness and secondness]. We may say that the bulk of what is actually done consists of secondness -- or better, secondness is the predominant character of what *has been* done . . . In general, we may say that *meanings* are inexhaustible.

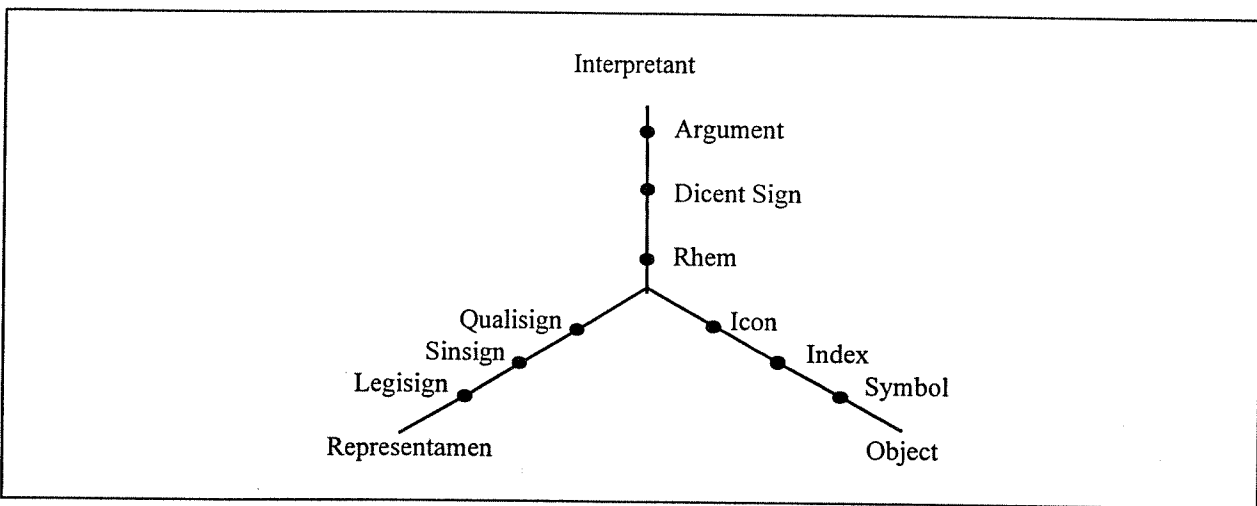
We are too apt to think that what one *means* to do and the *meaning* of a word are quite unrelated meanings of the word “meaning,” or that they are only connected by both referring to some operation of the mind.

Classes of signs

Before a classification of signs can be developed we need to divide the three aspects of the sign: the representamen, the object, and the interpretant, into three levels, these are called *trichotomies*. The trichotomies are based on the aforementioned phenomenology, viz. firstness, secondness and thirdness.

Each sign consists of *three* components: representamen, object and interpretant, which all have to be present to make a sign. The three components of the sign further have three elements, which reflects the three modes of being; firstness, secondness and thirdness. This trisection of the sign, and the further trisection of its components is essential to Peirce’s semiotics. The trisection of the trisection can be represented as:

Figure 4. Triadic Classification of Signs



The three inner categories, *rheme*, *icon* and *qualisign*, represent firstness, the middle categories, *dicent sign*, *index* and *sinsign*, represent secondness, and the outer categories, *argument*, *symbol* and *legisign*, represent thirdness.

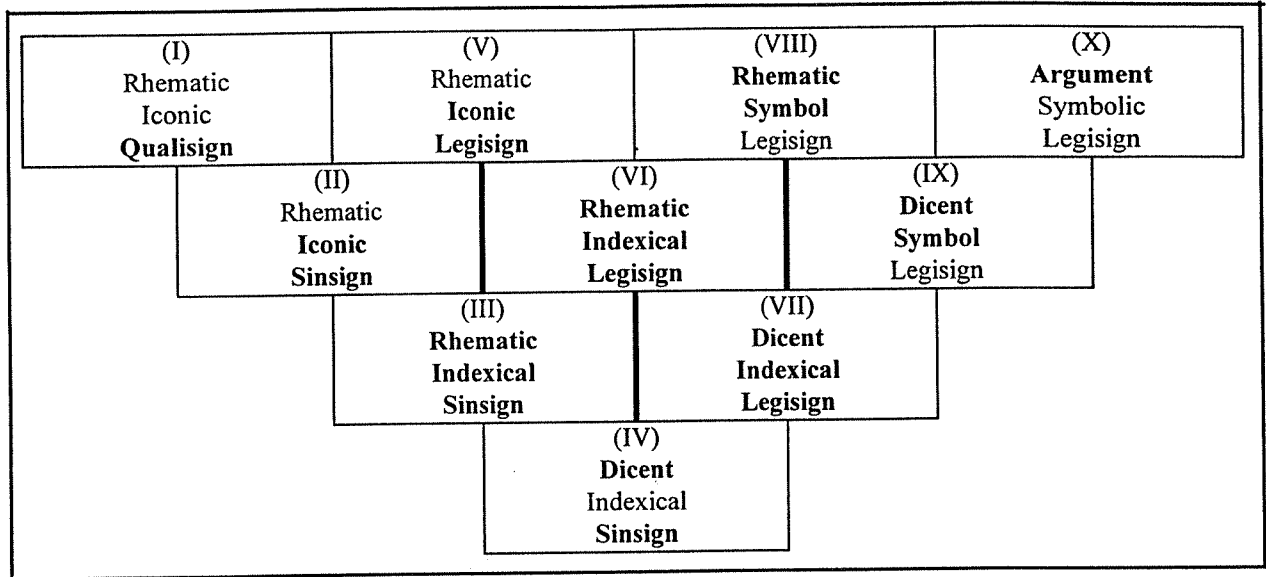
- The representamen is divided according to whether the sign itself is a mere quality (*qualisign*), an actual existent (*sinsign*), or is a convention (*legisign*). For instance the sign ‘A’ could be considered as 1) black lines, 2) a good example of the letter A, or 3) an expression of satisfaction with a term paper.
- The object is divided according to the sign’s relation to the object it represents. The sign could consist in having some character in itself (*icon*), in some existential relation to that object (*index*), or in its relation to the interpretant (*symbol*). An example of an icon is a pictogram where the sign resembles the object, a footprint which points to a person is an example of an index, and a symbol’s representation is based on the context in which it occurs.
- The interpretant represents the sign as a sign of possibility (*rheme*), a sign of fact (*dicent sign*), or a sign of reason (*argument*). Examples of these three are 1) nouns, 2) propositions, and 3) arguments, i.e. meaningful links of propositions.

The examples above are all rather weak since any sign is defined as a combination of the elements above, such that each sign consists of one element from each trichotomy. This is a further development of the aforementioned popular division of signs into three kinds.

By combining the elements Peirce defined ten categories of signs. Some possible signs are not defined since they are logically excluded. A *qualisign* will for instance always be a rhematic icon (because a mere quality cannot be a

convention). A symbol will always be a legisign (a symbol is a representation of its object based on the context, and a legisign is a sign based on conventions). An argument will always be a symbolic legisign (since an argument is thirdness in the interpretant and thereby requires a high degree of interpretation). The categories have been represented graphic as (Peirce, 1955, p. 118):

Figure 5. Ten categories of signs



To show the relationship between the categories of signs Peirce arranged the categories into ten boxes. The boxes are arranged according to their relationship. This relationship is based upon the number of aspects the categories share. Each pair of boxes next to each other share two aspects, e.g. the categories in boxes (I) and (II) share the aspects 'rhematic' and 'iconic', boxes (IV) and (VII) share the aspects 'dicent' and 'indexical', boxes (VIII) and (X) share the aspects 'symbol' and 'legisign'. Except for this rule are boxes separated by a thick line, i.e. boxes (II) and (VI), (VI) and (IX), (III) and (VII), these only share one aspect. The categories in boxes separated from each other by another box also only share one aspect. Boxes that are more than two boxes away from each other, e.g. boxes (I) and (IV), share no aspects.

THE CONCEPT OF SUBJECT

To distinguish the nature of the different elements and steps in the subject indexing process, we will here categorize the elements in the process according to Peirce's categorization of signs. By this categorization the difference between the elements, according to the level of individual interpretation which is required at each element is illustrated. This will clarify the difference between the elements according to how the each element refers to the meaning of what is represented.

Element 1 – the digital object

The digital object itself belongs to the complete thirdness, which means all the three aspects of the sign belong to *thirdness*. The digital object is therefore categorized as *argument* (category X). Peirce defined an argument as (Peirce, 1955, pp. 117-18):

An Argument is a sign whose interpretant represents its object as being ulterior through a law, namely, the law that the passage from all premises to such conclusions tends to the truth.

The digital object itself is the representamen, it gives meaning through laws, or conventions, which are 'determined' by humans in a cultural setting. The digital object could have been created differently and still have given the same meaning, or referred to the same object or idea. The digital object will therefore be a *legisign*. Legisigns are characterized by referring to their objects through a general law (or conventions). The connection between the digital object and its idea will take place through an interpretation. The interpretation is based on social and cultural conventions, therefore is the digital object is a *symbol*.

Element 2 – the subject

The subject, which here is defined as the understanding of the digital object perceived by the indexer, belongs to the category of *dicent symbols*. Dicent symbols are slightly different from *arguments*. Where the interpretant in arguments, is thirdness it is secondness in *dicent symbols*. This means that there is a lower level of interpretation needed. The object is still a symbol, which mean that the representation of the object depends on the context.

Element 3 – the subject description

The subject description is a formalization of the subject reached at element two. The subject description can be categorized as a *dicent indexical legisign*. Dicent indexical legisigns are characterized by directly pointing their objects out (since they are indexes), the subject description points to the subject by a precise description of it. Dicent indexical legisigns further give some information about their objects, since they are dicent signs they require some interpretation.

Element 4 – the subject entry

The subject entry is categorized as a *rhematic indexical legisign*. The subject entry points out its subject, and is therefore an index, it requires a minimum of interpretation, and is therefore a rheme. The subject entry therefore gives very little information about its object but merely points it out.

Summary

This analysis is based on the assumption that every element in the subject indexing process could be regarded a sign. The most complex of Peirce's categories is the category of arguments. A digital object is in this framework the most complex statement made, we therefore categorize digital objects as arguments. The next step in the subject indexing process is the subject. We regard this as less complex than the digital object, and therefore change the sign's relation to the interpretant from a sign of reason (argument) to a sign of fact (a dicent sign). Arguments and dicent symbols are legisign and symbols, since both are based on conventions (legisign), and they relate to their objects through on the basis of the sign's context (symbols). The subject description is also based on conventions (legisign) and a sign of fact (dicent sign), but there is less interpretation involved with the subject description, the trichotomy for the object moves to become index. This means that the subject description points to its object more directly than the two former elements. The last element, the subject entry, is as **the three previous** based on conventions (legisign) and like the subject description is points out its object. The trichotomy for the interpretant moves to become firstness (rheme), which means that the level of interpretation is on a minimum.

CONCLUDING REMARKS

An information retrieval system for retrieval of digital objects needs to represent the intellectual content of the digital objects to provide access to the digital objects based on the their meanings. The process of representing the intellectual content, or the subject, of digital objects takes a number of steps. The exact actions taken at each step have not been described in the literature, although some attempts have been made (e.g. Wilson, 1968).

When the subject indexing process is analyzed with the concept of unlimited semiosis the high degree of individuality of the process is stressed. It is shown how later elements of the subject indexing process depends on earlier steps taken in the process, and how the referent changes as the process takes place.

Each element of the subject indexing process belongs to different categories of signs. Each element requires different levels of interpretation and refers to its object differently. The first element of the process requires a high level of interpretation and refers to its object through conventions, whereas later elements requires minimal interpretation and refers to their objects by pointing to these objects.

A framework for studying the subject indexing process has been provided. This framework argues that each element is considered a sign and that each step is an interpretation of former elements in the process. It is argued that later steps in the process require less interpretation than former steps, and hence that the former steps are more critical for the result of the process than later steps.

Further study on the first steps of the process is critically needed. Especially is study on the analysis of the digital object needed, this, the first step in the process, is critical for the result of the subject indexing process.

NOTES

1 Peirce used 'sign' both to denote the primary sign (the physical entity we see) and the sign relation (the triadic relation between the three units in the Y-leg model).

2 Peirce names his phenomenology Phaneroscopy (from Greek *faneros*: evident, distinct).

REFERENCES

Bates, M. (1986). Subject access in online catalogs: A design model. *Journal of the American Society for Information Science*, 37(6), 357-76.

Blair, D.C. (1990). *Language and representation in information retrieval*. New York: Elsevier Science Publishers.

Brier, S. (1996). Cybersemiotics: A new interdisciplinary development applied to the problems of knowledge organization and document retrieval in information science, *Journal of Documentation*, 52 (3), 296-344.

Chu, C.M., & O'Brien, A. (1993). Subject analysis: the first critical stages in indexing. *Journal of Information Science*, 19, 439-54.

Dines Johansen, J. (1985). Prolegomena to a semiotic theory of text interpretation. *Semiotica*, 57(3/4), 225-88.

Eco, U. (1990). Almen semiotik og sprogfilosofi. (Universal semiotics and philosophy of language) *Almen Semiotik*, 1(1), 12-33.

Farrow, J.F. (1991). A cognitive process model of document indexing. *Journal of Documentation*, 47(2), 149-166.

Frohmann, B. (1990). Rules of indexing: A critique of mentalism in information retrieval theory. *Journal of Documentation*, 46 (2), 81-101.

Gluck, M. (1996). "Making sense of semiotics: Privileging respondents in revealing contextual geographic syntactic and semantic codes". Presented at Information Seeking in Context. A Conference on Research in Information Needs, Seeking and Use in Different Contexts. 14-16 August, at Department of Information Studies, University of Tampere, Finland.

Langridge, D.W. (1989). *Subject analysis: principles and procedures*. London: Bowker-Saur.

Miksa, F. (1983). *The subject in the dictionary catalog from Cutter to the present*. Chicago: American Library Association.

Peirce, C.S. (1955). *Philosophical writings of Peirce* (Selected and edited with an introduction by Justus Buchler). New York: Dover Publications.

Peirce, C.S. (1958). *Charles S. Peirce: selected writings: values in a universe of chance* (Edited with an introduction and notes by Philip P. Wiener). New York: Dover Publications.

Petersen, T. (1994). "Introduction". In T. Petersen, & P.J. Barnett (eds.), *Guide to Indexing and Cataloging with the Arts and Architecture Thesaurus* (pp. xiii-xvi). New York: Oxford University Press.

Taylor, Arlene G. (1994). 'Books and Other Bibliographic Materials'. In T. Petersen, & P.J. Barnett (eds.), *Guide to Indexing and Cataloging with the Arts and Architecture Thesaurus* (pp. 101-19). New York: Oxford University Press.

Warner, J. (1990). Semiotics, information science, documents and computers. *Journal of Documentation*, 46(1), 16-32.

Wilson, P. (1968). *Two kinds of power*. Berkeley: University of California Press.