A Framework for the Description of Generalised Documents

Roland Hjerppe, LIBLAB, Linköping University, Sweden

Abstract: Based on a generalisation of reading and texts the concept of document is generalised, to encompass all the types of objects collected by the memory institutions of societies. The description of such generalised documents and their contents is discussed in terms of description in general and as used in the memory institutions, noting the distinctions between describing single documents and contents (by themselves) and documents/contents among other documents/contents.

1. Introduction

The proposed framework for the description of generalised documents and their contents is a result of an investigation into similarities and differences in the cataloguing performed in the memory institutions of societies - libraries, archives, museums, heritage (monuments and sites) institutions, and aquaria and arboreta, zoological and botanical gardens. One of the outcomes of that work has been the insight that although comparative surveys of existing and planned cataloguing practice in these types of institutions, especially in terms of record structures and contents, seem to show that there are many similar descriptive elements, and that hence the same type of information system might be usable and that even merging of records could be technically feasible, with some adjustments, the fundamental philosophies of the institutions, the nature of their collections and items, and the uses of these are different enough to warrant a deeper investigation of what is common to them, at what level, and especially of the nature of their collections and of documents in general.

Emphasising other aspects these institutions might be called collecting institutions. David Bearman has in a report (Bearman, 1987) called them cultural repositories, accentuating the cultural properties of the collections and of the items (objects) in them. The communicative and retentive functions of these institutions and their collections are focused in this paper by assigning the label generalised document to cover all the items in all these institutions. The label "generalised document" is, however, not confined in this paper to the objects in these collections only but applicable in a wide variety of situations to objects that communicate, in particular to multimedia and the subjects of such presentations.

2. The concept of *document*

A document is, according to The Shorter Oxford English Dictionary, i.a.: "That which serves to show or prove something; evidence, proof. ... Something written, inscribed, etc., which furnishes evidence or information upon any subject, as a manuscript, title deed, coin, etc." Taking the cue from (Gerhart, 1987), who discusses a generalisation of the concept of text, a new definition is proposed:

1 supported by a grant from FRN - Swedish Council for Planning and Coordination of Research
a *document* is something that contains/carry a *text*.  

This definition shifts the interest from the object, the carrier, to the content, the text of the document. There is, however, now instead a need for defining *texts*. Rather than trying to define *texts* in terms of characters, or alphabets etc., which would be fraught with problems when we leave the realm of representations of ordinary languages, and which would move the attention to the (re)presentation, the uses of texts (i.e. texts are for reading) will be used to define *text*, hence:

a *text* is something that can be *read*.

By shifting the consideration, from texts as expressions, written by someone in some writing system, to the act of reading it becomes possible to regard e.g. the results of natural causes as *texts* since they can be *read* by properly trained *readers*. The act of *reading*, which still needs definition, is then defined as follows:

*reading* is an acquired skill

which implies in the reader an acquired competence to i.a.
- recognise a *document*,
- isolate its *text*,
- discern the representation of the *text* (and know its syntax etc.),
- comprehend the *language* of the *text*,
- understand the *surface meaning* of the *text*, and
- interpret, relate the *text* to oneself, to other *texts*, and to the world.  

The partial circularity in this last definition is by design. Implicit is also verifiability, consensus among *readers* on the *text*. The undefined terms will (for lack of space) be left undefined here since the aim of this paper is to discuss the description of *documents* and their contents. It should also be noted that in the definitions given *reading* is not restricted to the use of vision but is applicable to any of the senses, or combinations of them.

Every object that we attend to has a probably unlimited number of contents, depending on our interest in them and the context of our attention, etc., but with the definitions above the content (of interest) of a *document* is its *text*. In this approach to *documents* and their *texts* the *texts* are not resident in the *documents* but the result of a *reading* process. The *text* might thus be regarded as the result of a performance with the *reader* as performer and the *text* as a *score*. The performance analogy should, though, be taken as exactly that. The present discussion is furthermore delimited to *documents* that have some material manifestation.

The approach to *documents* and *texts* proposed above is an attempt to reconcile the hermeneutic emphasis of Buber on man as the essential dwelling place of language, as well

---

2To avoid repeating the adjective generalized in the rest of the paper the convention that something that is enclosed in asterisks *...* should be read as generalized ... (viz. *text* should be read: generalized text).

3These are only some of the demands that reading a text makes on our minds and bodies. See the chapter on Media as Epistemology in (Postman, 1985,) for a simple account, or any textbook on cognitive psychology e.g. (Best, 1992).
as his recognition of works of art as *texts*, to be *read* and interpreted, elucidated in (Kepnes, 1992), with Goodman's theory of symbols oriented discussions of art in (Goodman, 1976), and especially his reasoning on worldmaking in (Goodman, 1978). The 'romantic epistemology' of (Novitz, 1987), which focuses fiction as a source of knowl-
edge, even about the world, has similarly been extended to encompass *documents*.

One of the main uses of *documents* is in the process of generating new ones. Many *texts* are thus *read* before and concurrently with the creation of new *texts*. The unfinished *document* (the draft, sketch, prototype, ...) has in these cases a special status since its *text* is being both *read* and *written* by the author/creator during its generation. In Figure 1 below is a schematic model of the employment (*reading* as well as *writing*) of the *texts* of *documents*. In this model the *reader* and the *writer* have, mainly for clarification of functions, both been divided into a perceiver/reader + reader/constructor, and an operator/agent + agent/articulator respectively.

Figure 1. Factors in the employment of *documents*
A partly implicit assumption in the exposition so far has been that *documents* and their *texts* are simple, i.e. that they are of a single modality. This assumption is partly a result of our conceiving texts as single media. There can in the case of *documents* be no prior assumption of which, or the number of, senses that are engaged in the *reading*. Neither can the *documents* be assumed to be static, immutable, quiescent objects, waiting to be *read*. They might also be autonomous agents or subjects. *Documents* can, furthermore, be compounds, composed of parts that in themselves are *documents*.

So far the question of meaning; why *documents* are employed, what the "residual" effect of the employment of *documents* are in the *reader/writer*, have been dodged. In the list above of acquired competencies implied in the acquired skill of *reading*, and in Figure 1. above, there are mentions of "surface meaning" and of "interpretation" that need further explication.

One of the extreme consequences of the stance taken is of course that the world around us is a *document*, to be *read*, as a whole and in details. The world around us is the first *document*, learning the *reading* of which, as a social activity, enables us to learn reading, which provides us with access to texts that enable us learn *reading*. The skill of *reading* is initially acquired from other *readers* and the employment of *documents* is mainly a social activity. The implications of this is that

the meaning of a *text* is a social construct,

with two components, the public and the private meaning. The public meaning is that which earlier in this paper has been labelled surface meaning, and which is verifiable by consensus among several *readers*. The private meaning is the result of an interpretation that relates the *text* to the *reader*, to other *texts*, and to the world. These two components are of course closely interrelated.

3. Description

A description of a *document* is an expression of a *reading* of that *document* that concerns the *document*, its *text*, or its meaning. The descriptions that are of interest here are those that themselves are intended as parts of *documents* - catalogues - which therefore can be regarded as secondary *documents*, and which are used mainly as substitutes for the original, primary *documents*.

In the general case there is a need to distinguish two situations in the description of *documents*.

- the *document*/*text*/meaning among other *documents*/*texts*/meanings
- the *document*/*text*/meaning by itself

The first situation is the one dominant in the memory institutions. The objective of description can in these cases be regarded as relation and identification, based on two primary, associated types of categorisation: universals and particulars, the samenesses and their differences. In terms of information systems the concerns are to establish the sets of
attributes that are applicable to all of the items in a collection and the values of which, in each specific combination, enables a differentiation of that item from the other items in the collection. *Document* description in the memory institutions seems thus to conform well with Bertrand Russell's theory of description.

Figure 2. The descriptive process for items of collections

Figure 2 above is hence intended to show how in the description of an item there must first be a decision on what item class, of a collection, that item is part of, and consequently which item class description should be used as a template for the description of the item,
and that there is an interdependence in that only by going through this process can the classes as well as their descriptions, and hence the item descriptions be established.

This descriptive principle seems to be valid mainly for *documents*, i.e. for what often is called formal description. The attribute values are either names or largely based on what Lakoff (Lakoff, 1987) calls basic-level categories. There is very little of abstraction or specialisation among the attribute values. The descriptions are mainly aimed at identifying, in two senses: separating each *document* from other (similar) *documents*, and enabling identification by comparing the description and the *documents*. The *texts*, are to a large extent not described by the memory institutions.

In this process model the interest is in the items and the resulting descriptions, not in what happens inside the *reader* of the items/*documents*. In (Moustgaard, 1981) there is an interesting attempt to analyse the descriptive process from a psychological, inside, perspective that in many ways is a complement to the present paper.

The descriptions of the public meaning, usually called content or subject description, are largely based on systems of abstract categories, abstractions or specialisations of basic-level categories - systems for knowledge organisation. The aim is categorisation in terms of assigning the *document* to a class of (similar) *documents*. One of the major differences between the memory institutions in this respect is that whereas in libraries, especially those concerned with scientific research, the description of the meaning (subject) of a text is related to a structuring that mirrors the current (at the time of description) state of disciplines etc. this is not directly the case for the other memory institutions. In other words, the description of the meaning of an item in e.g. a museum or an archive does not directly relate it to disciplines etc. but instead to other items, other *documents* or to statements about the world..

The second situation, the *document*/*text*/meaning by itself, is the one prevalent in e.g. criticism and analysis of art, literature etc. It is of course not possible to completely isolate any *document*/*text*/meaning from other ones, but the emphasis is usually on a single *text* or a small number, usually by the same creator. The descriptions are concerned with both public and private meaning and the interdependencies between meaning and *text* (form and content), and their relations to specific other single *texts* and meanings, and the world. The *document* aspect is mostly neglected. The descriptions are in the second situations often discursive and they are also usually extensive *documents*, with the difference that criticisms/analyses usually are mainly textual.

The descriptive expressions, the *descriptors*, used in the first situation - catalogues - can be extractions, derivations, or assignments. Extractions are parts taken from the *document*, as is, e.g. excerpts. Derivations are the results of some kind of processing, of the *text*, or parts of it. Derivations can be rule-based transformations, or mappings onto a system of knowledge organisation, external to the *document*. Assignments, which arise in the *reader* during the *reading*, can be ad-hoc, or systematic.

The *descriptors* can additionally be distinguished as to whether they describe the *document*, or the relations of the *document* to other *documents*, in the same collection or several collections. In Figure 3 below is shown schematically some of the
different types of *descriptors* and their relations to *documents*, collections, and systems of knowledge organisation.

4. Why describe?

The principal considerations for the design and of the descriptions in catalogues have usually been conciseness and potential for arrangement in sequences. Descriptions have therefore usually been in the form of formalised texts. That is partly a matter of technology and partly of convention. It is a matter of technology in that other and fuller forms of representation have been difficult to employ and a matter of convention in the sense that descriptions that employ other forms of expression than text, that hence are *documents*, have been regarded as such. When it becomes possible to access and use *documents* as easily as we now use catalogues then the question of why and how to describe will need
renewed attention. To the extent that descriptions are result of interpretations and add value to the *documents* or the collections of *documents* will there be benefit in catalogues as we know them today. Formal description will, however, partly be inherent in the explicit structure of the object oriented *documents* of tomorrow.

References:


Gerhart, Russell AM (1987) A Generalized Conception Of Text Applied To Both Scientific And Religious Objects. ZYGON v.22 n3 pp.299-3161


