

Syracuse University

SURFACE

School of Information Studies - Faculty
Scholarship

School of Information Studies (iSchool)

1-1-1991

The Importance of Factors that are not Document Attributes in the Organization of Personal Documents.

Barbara H. Kwasnik
Syracuse University

Follow this and additional works at: <https://surface.syr.edu/istpub>



Part of the [Library and Information Science Commons](#)

Recommended Citation

Kwasnik, Barbara H., "The Importance of Factors that are not Document Attributes in the Organization of Personal Documents." (1991). *School of Information Studies - Faculty Scholarship*. 146.
<https://surface.syr.edu/istpub/146>

This Article is brought to you for free and open access by the School of Information Studies (iSchool) at SURFACE. It has been accepted for inclusion in School of Information Studies - Faculty Scholarship by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

THE IMPORTANCE OF FACTORS THAT ARE NOT DOCUMENT ATTRIBUTES IN THE ORGANISATION OF PERSONAL DOCUMENTS

BARBARA H. KWAŚNIK

School of Information Studies, Syracuse University, Syracuse, NY 13244

The purpose of the study was to investigate and describe the way in which people organise documents in their offices. Eight university faculty members were asked to describe their own offices in terms of the organisation of documents. Each respondent was also asked to sort a typical day's mail. Following the data analysis, four of the eight respondents were interviewed again, at which time the researcher used the results of the analysis as a guide and tried to sort and classify each respondent's mail as he or she might have. An inductively created content analysis of the data was performed and the results suggest that the criteria or factors that people take into account when classifying documents consist not only of document factors (such as the document's topic or form) but also, to a very important degree, of situational factors (such as the use to which the document is to be put). In addition, it was shown that people are able to articulate the process by which classification decisions were made, and the data produced by this articulation lend themselves to analysis at a level which can yield general rules about the behaviour. One implication of this study is that, in designing systems for organising documents, it might be advantageous to explore ways of modeling typical contexts of document use, since these seem to be very important in classification decisions made within personal information environments.

INTRODUCTION

This paper summarises an investigation of the process by which people organise and classify documents in their own personal information space [1]. In particular, the research attempted to describe one aspect of classificatory decisions, namely, those important factors that are not document attributes. These factors are referred to as 'context', this being defined as a person's goals, purposes, knowledge, history, predisposition, constraints and understandings of the situation in which the person finds him or herself on the occasion of making a classificatory decision.

The main objectives of the study were to:

1. elicit a description of context;
2. identify what material phenomena (documents) are significant in this context;
3. observe and describe how and under what circumstances the documents are differentiated, classed, and integrated;
4. isolate from this process those dimensions of the user's context and features of documents that are salient for the user; and
5. identify patterns, or 'enduring reciprocities' [2] of this behaviour.

The work is indebted to several researchers who have investigated the organisation of personal collections of documents [3-6]. This investigation extends their work to include a consideration of not only the documents themselves and their clustering or grouping, but also the importance of the situation in which the activity takes place.

BACKGROUND

The importance of studying behaviour in the context in which it occurs was emphasised [2, 7]. Individual perspectives and varying contexts, as well as the fact that objects are not well defined require that the description be multidimensional and that 'things' be seen not only with respect to each other, but also with respect to the situation of the classifier [8]. Things and events are not mutually exclusive, that is, they can belong to more than one category at a time, and can belong to categories 'more or less' [9-11]. Furthermore, the principle of transitivity is not uniform among people [12]. Finally, varied perspectives influence the scale, that is, the level of attention at which an object is perceived [13].

STUDY DESIGN AND PROCEDURES

A pilot study was conducted to test the main concerns of methodology which were:

1. the difficulties of collecting verbal data;
2. the feasibility of collecting data in as natural a setting as possible;
3. the feasibility of using open interviews and 'thinking out loud' protocols to allow subjects to generate their own descriptions, labels, and relationships;
4. the ability of subjects to co-operate in terms of verbalising and offering information with respect to context and process.

The results of the pilot study supported the study design.

Sample. The participants for eight case studies were recruited from university faculty by means of personal introduction. The sample was selected to include individuals representing a variety of characteristics that might influence classificatory decisions: academic discipline, sex, length of time in the university and a variety of working environments. Although comparisons were not made along these criteria, the aim was to have a reasonably broad range of participants.

Procedures. The study consisted of several stages of data collection and analysis. Interviews, observation, and thinking-out-loud protocols were the main data collection techniques employed. The interview sessions were scheduled in the following way:

Session 1. On the first visit, the participant was asked to provide a 'guided tour' of the office or of any other space in which materials were stored and used, describing the various piles, drawers, shelves, desktop, and so on. This procedure was repeated until all materials were covered. The resulting interviews were tape-recorded and transcribed.

Session 2. The participant was asked to sort a day's mail following the usual procedure as closely as possible. He or she was asked to 'think out loud' while sorting the mail, describing each piece and specifying what would be done with it. This protocol was recorded and transcribed.

Session 3. Four of the eight subjects were asked to save a few days' worth of mail and a few days' worth of documents that had been used recently. These items were to be placed on a pile and not disposed of as they might have been under usual circumstances. The researcher, using the descriptions and rules generated by the data analysis as a guide, then attempted to sort the pile of documents in the same way the subject might have done. The subject was asked to comment on the accuracy of the researcher's decisions, and if they were wrong, to comment on the reason for the error. These interviews were tape-recorded as a memory aid but not transcribed.

ANALYSIS AND DISCUSSION

The output of analysis can be divided into three groups, each group contributing to the overall picture of a person's classificatory decisions within the context of an academic office.

The identification of important dimensions. The analysis of the data showed that documents are identified and classification choices are made in situations that can be described by a variety of dimensions. One of the aims of this study was to determine which dimensions of description and classification are most important, that is, which ones determine the primary 'meaning' of a document, and then to consider the occurrence and frequency of these dimensions.

Important dimensions were extracted from the interview and thinking-out loud protocols using the following procedures: first, each instance in which a participant identified a 'document' (using his or her own judgment to define a document) was marked. In speaking of documents, participants often used modifying phrases indicating the circumstances of the classificatory decision. These modifying phrases were interpreted by the researcher and summarised by brief terms or labels. The labels were then defined into coding categories.

<i>Situation attributes</i>	<i>Order/scheme</i>
Access	Accumulation
Circumstance	Arrangement
Need/Requirement	Group
Ownership of the document	Separate
Related to me	Unfinished arrangement
Room/Space	
Source	<i>Time</i>
Use/Purpose	
<i>Document attributes</i>	<i>Value</i>
Author	Important
Form	Interesting
Topic	Needs improvement
Title	Not valuable
Physical attributes	Secret/Confidential
	Unspecified value
	Works for me
<i>Disposition</i>	<i>Cognitive state</i>
Change	Don't know
Discard	Want to remember
Keep	'Just know'
Locate	
Postpone	

FIGURE 1. *Descriptive coding categories that represent dimensions used in making classificatory decisions*

after multiple passes through the data, and a codebook was built up iteratively. Finally, the entire corpus was coded using these coding categories. Each coding category represented a dimension along which classificatory decisions were made.

For example, the following two instances of classificatory decisions: on the top shelf are books that are very seldom used correspondence I must deal with immediately goes into my briefcase can be described by the same set of coding categories:

on the top shelf- LOCATION - into my briefcase
 books - FORM - correspondence
 very seldom - TIME - immediately
 used - USE - deal with

TABLE 1. Frequency of descriptive dimensions used in making classificatory decisions

Dimensions	Session 1 'Grand Tour'		Session 2 'Mailsort'		Total		
	No.	%	No.	%	No.	%	
<i>Situation Attributes</i>							
Access	86	2.2	2	0.3	88	1.9	
Circumstance	300	7.7	47	7.3	347	7.6	
Need/Requirement	68	1.7	15	2.3	83	1.8	
Ownership of document	21	0.5	4	0.6	25	0.6	
Related to me	111	2.8	15	2.3	126	2.8	
Room/Space	56	1.4	2	0.3	58	1.3	
Source	127	3.3	64	9.9	191	4.2	
Use/Purpose	516	13.2	77	11.9	593	13.1	
<i>Subtotal</i>	1,285	32.8	226	34.9	1,511	33.3	
<i>Document attributes</i>							
Author	66	1.7	3	0.5	69	1.5	
Form	576	14.8	93	14.4	669	14.7	
Topic	338	8.7	50	7.8	388	8.5	
Title	74	1.9	14	2.2	88	1.9	
Physical Attributes	113	2.9	12	1.9	125	2.8	
<i>Subtotal</i>	1,167	30.0	172	26.8	1,339	29.4	
<i>Disposition</i>							
Change	35	0.9	1	0.1	36	0.8	
Discard	34	0.9	21	3.3	55	1.2	
Keep	71	1.8	14	2.2	85	1.9	
Locate	291	7.5	79	12.2	370	8.1	
Postpone	19	0.5	10	1.6	29	0.6	
<i>Subtotal</i>	450	11.6	125	19.4	575	12.6	
<i>Order/scheme</i>							
Accumulation	33	0.8	3	0.5	36	0.8	
Arrangement	193	5.0	6	0.9	199	4.4	
Group	64	1.6	17	2.6	81	1.8	
Separate	43	1.1	7	1.1	50	1.1	
Unfinished order	36	0.9	1	0.2	37	0.8	
<i>Subtotal</i>	369	9.4	34	5.3	403	8.9	
<i>Time</i>	<i>Subtotal</i>	343	8.8	43	6.7	386	8.5
<i>Value</i>							
Important	24	0.6	8	1.2	32	0.7	
Interesting	36	0.9	9	1.4	45	1.0	
Need improvement	48	1.2	1	0.2	49	1.1	
Not valuable	32	0.8	12	1.9	44	1.0	
Unspecified value	38	1.0	5	0.8	43	0.9	
Secret/Confidential	3	0.1	1	0.2	4	0.1	
Works for me	32	0.8	1	0.2	33	0.7	
<i>Subtotal</i>	213	5.4	37	5.9	250	5.5	
<i>Cognitive state</i>							
Don't know	35	0.9	5	0.8	40	0.9	
Want to remember	13	0.3	2	0.3	15	0.3	
'Just know'	21	0.5	1	0.2	22	0.5	
<i>Subtotal</i>	69	1.7	8	1.3	77	1.7	
<i>Grand total</i>	3,896		645		4,541		

This analysis yielded an inventory of document labels and an inventory of dimensions, as well as an indication of how frequently each dimension was invoked with respect to the classification of documents. Once the entire corpus had been coded, it was possible to merge and rearrange the categories so that extremely fine levels of distinction that accounted for a very small proportion of the data were collapsed into more inclusive categories. This yielded thirty-three categories which were then classified by the researcher into seven even broader groups: situation attributes, document attributes, disposition, order/ scheme, time, value, and cognitive state. Figure 1 shows these thirty-three dimensions of classificatory decisions arranged into the seven broader categories.

Frequency of descriptive dimensions. Table 1 shows the frequency of descriptive dimensions for all cases. The dimensions have been divided into the main groups, and separate figures are shown for descriptive dimensions in Session-1 'Grand Tour' interviews and for those in Session-2 Mailsort, thinking-out-loud protocols. The percentage column shows percent of all dimensions accounted for by a given dimension under each interview condition.

The figures show that dimensions describing situation attributes account for the largest percentage of descriptive dimensions, followed by dimensions describing document attributes, followed by those describing how a document will be disposed of.

There are several aspects to the dimensions identified in the data as important to classificatory decisions. If frequency alone is considered, we can see that form, use, topic, location, circumstance, and time are among the most important criteria people consider in making a decision about a document. Clearly, it is necessary but not sufficient to consider document attributes alone, the mainstay of our traditional document classification systems. Document attributes need to be considered in conjunction with situational or contextual attributes in order to get an accurate description of classification decisions made by individuals for their own documents.

Rules. Each identified classificatory decision in all the cases was further analysed to discover and describe not only the frequency of classificatory dimensions but also the order in which they are invoked. Each classificatory decision was rephrased by the researcher as a rule. Whereas the previous section addressed labels of documents as evidence of the cognitive process of classifying them, the rules address the physical disposition of the documents - the actual placement of documents in physical containers or locations – also signifying a classificatory decision.

Each rule is composed of two parts. The first part, the IF statement, sets out the conditions for the classificatory decision. The second part, the THEN part, sets out the disposition of those documents when the specified conditions are present.

For example, the following excerpt of text (representing classificatory decisions made by one of the subjects) was reformulated into the rule that follows it. Creating a rule involves two steps: 1. analysing the text into its components, and 2. formulating these components into IF-THEN statements. In addition, each component is then described by a label representing a descriptive dimension of classificatory decisions.

The sample text: on the right hand side of the desk are file folders that are courses taught last year and not yet refiled. And also examinations from the last semester. So in other words all that stuff is nothing that's used on a daily basis or even a monthly basis.

The rule generated by this text:

	IF:	file folders for courses taught last year not refiled yet	Form Topic Use Time (tense) Unfinished Time
	OR		
	IF:	examinations from last semester	Form Time (age)
AND			
	IF:	not used daily or even monthly	Use Time (frequency)
THEN:		place on right side of desk in a pile	Locate Order

Each rule can be seen as an intersection of optional or necessary conditions for a certain classificatory decision.

Testing. What we really want to know as observers and describers of another person's cognitive process is whether our observation in fact reflects the way a person thinks under a given set of circumstances. Toward this end, a final

session was conducted with four of the eight participants to see if the descriptions and rules generated through the data analysis would hold up reasonably well in a situation similar to the one originally observed in Sessions 1 and 2. Each subject saved a few days' worth of mail and documents and the researcher attempted to sort them in the same way the person might.

Possible outcomes of the task attempted by the researcher in Session 3 were as follows:

1. the classification decision made by the researcher was correct
 - a. based on a literal match with previously provided information
 - b. the classification of a new or unfamiliar document was correct based on extrapolation from a general rule

TABLE 2. *Results of researcher's attempts to sort the documents of four subjects*

TABLE 2. *Results of researcher's attempts to sort the documents of four subjects*

					<i>Total</i>	
					<i>No</i>	<i>%</i>
<i>Correct decision</i>						
information provided previously	6	5	1	2	14	13
rule applied to new object	14	22	6	10	52	49
<i>Total correct</i>	20	27	7	12	66	62
<i>Partially correct decision</i>	5	8	4	5	22	21
<i>Incorrect decision</i>						
rule failure	2	2	1	1	6	6
insufficient information	5	3	0	4	12	11
<i>Total incorrect</i>	7	5	1	5	18	17
<i>Grand total</i>	32	40	12	22	106	100

2. the classification was partially correct and partially incorrect
3. the classification was incorrect
 - a. due to the failure of the rules, i.e. a failure of the analysis
 - b. due to insufficient or incorrect information, i.e. a failure of the data collection

In all, 106 attempts at classifying the subjects' mail were made. Table 2 shows the distribution of correct, partially correct and incorrect predictions.

Almost two-thirds of all the predictions were correct, and almost half of all were correct based on the appropriate application of a general rule. The greatest factor in rule failure was the failure to invoke the correct dimension, or failure to invoke it in the proper order. The greatest factor in incorrect decisions due to insufficient information was that of an intervening circumstance or event that occurred between the original interview and the present trial.

The results suggest that the description of people's classificatory behavior generated by the analysis of the data was capable of predicting behaviour, and that even when it did not, this was not because the description was incorrect but only that it was incomplete.

CONCLUSIONS AND IMPLICATIONS

The purpose of this study was to investigate and describe in a general way the process by which people organise and classify their own documents in their own personal information space. The findings of this study are based on only eight cases, and the classificatory decisions made by the eight subjects were explored at only one or two points in time. For this reason it is not possible to generalise the findings beyond the participants of this study. Nevertheless, the data strongly suggest that context, and not only the features of the objects being classified, plays an important part in the classificatory decisions.

Another finding is that, in this study, the relationship of context to behaviour remained relatively stable, that is, despite the variety of human behaviour, it was possible to describe patterns of those dimensions that play a part in classificatory decisions. The data suggest that important dimensions of classificatory decisions apply across cases, even when these cases were selected to represent a broad range of individuals from within the academic community. Even without further testing for the influence of individual differences, it is possible to identify some dimensions of classificatory behaviour that occur in all cases and in relatively stable proportions.

In general, the data suggest that for most decisions, there is an interaction of dimensions important to that decision. A person makes classification decisions within a context and for a purpose, but also within the constraints of physical objects and a physical environment. Thus, neither the document attributes nor the situation attributes can be considered independently.

These findings have implications for most conventional systems of classification, which are based almost entirely on the document's attributes - its form, its 'topic', its author and title. While such practices have been defended by the argument that it is not possible to incorporate situational attributes into these systems because (a) contextual attributes are not easily defined, and (b) contextual attributes differ too greatly from one person to another and from one situation to another, the findings of this study suggest that contextual attributes can, in fact, be described from what people tell us, that on a general level they remain fairly stable across situations, and most important, they are extremely important to individuals when they classify documents for themselves.

The study was designed to show ranges of dimensions, to demonstrate that such dimensions could be identified, and to suggest that patterns do exist. With a larger number of subjects and with a longer duration, the dynamic process of classification could be more completely described. What the data in this study suggest is that the process is, in fact, dynamic, that the range of dimensions is manageable and that patterns can be described with enough detail to emulate classificatory decisions on a general level.

REFERENCES

1. KWAŚNIK, B.H. *The influence of context on classificatory behaviour*. PhD thesis, School of Communication, Information, and Library Studies, Rutgers. The State University of New Jersey, 1989.
2. MISCHLER, E. G. Meaning in context: is there any other kind? *Harvard Educational Review*, 49(1), 1979, 1-19.
3. CASE, D. O. Collection and organization of written information by social scientists and humanists: a review and exploratory study. *Journal of Information Science*, 72(3), 1986, 97-104.
4. JAHODA, G. *Information storage and retrieval systems for individual researchers*. New York: Wiley, 1970.
5. JAHODA, G., HUTCHINS, R. D. and GALFORD, R. R. Characteristics of use and personal indexes maintained by scientists and engineers in one university. *American Documentation*, 17, 1966, 71-75.
6. MALONE, T. W. How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Office Information Systems*, 7(1), 1983,99-112.
7. TYLER, S. A. Introduction. In: TYLER, S. A., ed. *Cognitive anthropology*. New York: Holt, Rinehart & Winston, 1969.
8. BROOKES, B. C. Objective maps and subjective landscapes: the foundations of information science. Part III. Quantitative aspects. *Journal of Information Science*, 2, 1980, 269-325.
9. MERVIS, C. B. and ROSCH, E. Categorization of natural objects. *Annual Review of Psychology*, 32, 1981, 89-115.
10. ROSCH, E. Cognitive reference points. *Cognitive Psychology*, 7(4), 1975, 532-547.
11. ROSCH, E., MERVIS, C. B., GRAY, W. D., JOHNSON, D. M. and BOYES-BRAEM, P. Basic objects in natural categories. *Cognitive Psychology*, 8, 1976, 382-439.
12. HAMPTON, J.A. A demonstration of intransitivity in natural categories. *Cognition*, 12, 1982, 151-164.
13. JONES, K. P. The environment of classification: the concept of mutual exclusivity. *Journal of the American Society for Information Science*, 24(2), 1973, 157-163. (Revised version received 8 April 1991)