60 YEARS OF THE BEST IN INFORMATION RESEARCH

Some thoughts on classification for retrieval

Karen Spärck Jones
University Mathematical Laboratory, Cambridge, UK

Abstract

Purpose – This paper, originally published in 1970, considered the suggestion that classifications for retrieval should be constructed automatically and raised some serious problems concerning the sorts of classification which were required, and the way in which formal classification theories should be exploited, given that a retrieval classification is required for a purpose. These difficulties had not been sufficiently considered, and the paper, therefore, aims to attempt an analysis of them, though no solutions of immediate application could be suggested.

Design/methodology/approach – Starting with the illustrative proposition that a polythetic, multiple, unordered classification is required in automatic thesaurus construction, this is considered in the context of classification in general, where eight sorts of classification can be distinguished, each covering a range of class definitions and class-finding algorithms.

Findings – Since there is generally no natural or best classification of a set of objects as such, the evaluation of alternative classifications requires either formal criteria of goodness of fit, or, if a classification is required for a purpose, a precise statement of that purpose. In any case a substantive theory of classification is needed, which does not exist; and, since sufficiently precise specifications of retrieval requirements are also lacking, the only currently available approach to automatic classification experiments for information retrieval is to do enough of them.

Originality/value – Gives insights into the classification of material for information retrieval.

Keywords Information retrieval, Classification

Paper type Conceptual paper

It is a commonplace of documentation and information retrieval that classification should be involved at some point in the document description, storage, and retrieval process. We may want either to classify documents to simplify searching, or we may want to classify an initial indexing vocabulary to provide a set of document descriptors, in this case to optimize request document matching. More sophisticated approaches involve request classification as well.

Now while the idea of classification is not a new one in this context, the suggestion that document or vocabulary classifications should be constructed automatically has raised new problems. Purely practical problems of scale arise, but the fact that a computer functions as a black box in forming a classification brings up some more interesting questions. One is what sort of classification we should look for. The mere fact that the human classifier observes his classification as he constructs it, so that he

This paper was previously published in *Journal of Documentation*, Vol. 26 No. 2 (1970), pp. 89-101. It has been included in this issue as part of a series of articles celebrating 60 years of the best in information research in *Journal of Documentation*.

Journal of Documentation
Vol. 61 No. 5, 2005
pp. 571-581

0022-0418
DOI 10.1108/00220410510625796

necessarily finds it plausible, may mean that he does not examine the principles on which it is based in a sufficiently critical way, or even that he does not formulate them properly or apply them consistently. Automatic classification can cause rude shocks, in that apparently satisfactory initial grouping principles can lead to a quite unexpected final output, when their implementation by a formal procedure precludes modifications and omissions during the grouping process. The second important question is how we should exploit classification theory, given that we have a specific purpose, namely that of retrieving documents relevant to requests. This paper does not supply any answers to these questions; but the difficulties which arise in organizing experimental research in this area, and in correlating the results of different projects, have suggested that something can be gained simply from a discussion of the points raised by the questions.

Suppose, therefore, that we imagine we are interested in constructing a retrieval thesaurus automatically, i.e. that we want groups of keywords so that the occurrence of one word in a request is matched by the occurrence of its classmates in documents. The assumption is that if the classes are correctly constructed this will lead to the retrieval of documents which are about the same thing as the request, though this is expressed in different words. This is quite a straightforward idea: I am using it here primarily as a peg to hang the general discussion on. In what follows, therefore, I shall take word classification as an example, but all the main points apply equally to document classification, though there are differences of detail.

Now if we say that we want to produce a keyword classification this leads us naturally to some such argument as the following concerning the sort of classification we require. We want to group keywords so that we obtain sets of words which are substitutible for one another, i.e. so that one member of a class in a request can be matched by another in a document. Synonyms are obvious examples of substitutible terms. If we want to construct our classification automatically, however, we cannot inspect the meanings of words to see if they are substitutible; we are obliged to look for some other information about verbal relationships which can be treated as pointing to substitutibility, and which is amenable to machine manipulation. The usual assumption here is that this is provided by facts about the occurrences and co-occurrences of words in documents. For if two words always co-occur they are clearly substitutible, because either can be used to retrieve the same documents; and while two words rarely co-occur to this extent, it is an empirical fact that words tend to co-occur in documents. We can then say that words which tend to co-occur can be treated as if they always co-occur, i.e. can be regarded as substitutible. Generalizing from this gives us sets of words which tend to co-occur, where some words do not necessarily co-occur themselves, but co-occur with common words: this is true of synonyms, for instance. We can then say that what we are interested in are topic classes of words which tend to appear in the same subject contexts, where these can hopefully be identified given initial information about the occurrences of the words in a collection of documents. The important point here is that this kind of information can clearly be manipulated automatically very easily.

I shall not go into this argument in more detail here, or seek to justify it; for the present purpose it is sufficient that it is usually proposed as a base for automatic thesaurus construction. What is more important is that if we adopt the argument, it follows that we are interested in a certain sort of classification, namely a polythetic and multiple one. By this I mean a classification in which the members of a class do not necessarily all share one or more common properties, and in which individual items can appear in more than one class. The reasons for this conclusion are obvious enough. Thus we are most unlikely to find sets of words which all occur in the same document or documents; looking for sets of words which tend to co-occur, which is all we are likely to find in practice, means looking for sets of words which share sets of documents, like *a*, *b*, and *c,* for example, if *a* occurs in documents 1 and 2, *b* in 1 and 3, and *c* in 2 and 3. This is a natural consequence of the fact that the documents in a collection, though they may be topically related, are not likely to be identical in both subject matter and vocabulary. Again, allowing words to appear in more than one class reflects not only the fact that even technical words may have different meanings within a specialized collection, so that they should have different substitutes, but also the fact that a word may be used in the same sense in different contexts, where its different topical affiliations will be represented by different co-occurence patterns and should again be reflected by different sets of substitutes.

In addition, if we also think we have no reason for extending the range of substitutes for a given term too far – and there is independent experimental evidence to support this view – we are not interested in an ordered classification. An ordered classification is one in which there are systematic relationships between classes which can be exploited to increase the number of possible substitutes for a given term. But it seems to be the case that while we want to allow some substitution, we do not want to allow too many alternatives for a word. We may indeed go further and say that we want an *un*ordered classification, because we might well get unsatisfactory results if we constructed an ordered classification and then disregarded its structure.

We thus arrive at a general statement about the sort of classification we want, that is, about its overall formal properties: we can say that we are interested in polythetic, multiple, unordered classifications. Now when we add to this the fact that we are typically interested in classification on a large scale, where this means that we may want to classify several hundreds or even thousands of keywords on the basis of information obtained from several thousands or even tens of thousands of documents, it will be apparent that we are faced with a serious problem. This is that we are interested in classifications which are particularly complex in character and at the same time are trying to discover them in large sets of objects. The scale difficulty is not, however, the most important one, though it is awkward enough in practice and is bound to influence the choice of algorithm for forming a classification. The real difficulties are connected:

(1) with the sort of classification we are interested in; and

(2) with classification in general.

Thus under the first head, any attempt to construct a classification of the sort in question simply reveals the lack of suitable classification methods, and further, of programmable algorithms for finding classes on a given definition. Under the second head we encounter what may be called the evaluation problem which is associated with classification in any guise but which is emphasized, as we saw earlier, by the use of computers: this is how we decide whether a classification is a good one, or which is the best of several classifications.

Now we may say that while there is a lack of suitable classification methods to choose from, it is not true that there are none. Needham's (1963) theory of clumps was originally intended to meet our retrieval requirements (Sparck Jones, n.d.). But the range of alternatives is not large, and the status of the methods in question and their relation to one another is not clear. So we in fact have no obvious choice. Equally, we can say that in information retrieval we have a means of evaluating a classification, namely whether the results are satisfactory when standard retrieval performance measures, like recall and precision ratios, are used. But quite apart from the problem of setting up really satisfactory retrieval performance measures, the relationship between the purpose for which a keyword classification is required and the technique adopted to generate it is far from clear. The same applies to document classification. The attempt to construct retrieval classifications automatically thus raises quite general questions about classification: and in the discussion which follows these will be considered in the hope that some guidance as to the best approach to retrieval classification can be obtained.

Suppose, therefore, that we start by looking briefly at classification in a wider context, by contrasting it with some other kinds of data processing which do not really produce classifications though they are sometimes treated as if they do, or are exploited as a base for classification. We can then contrast the sort of classification we are interested in with other sorts.

We can describe classification in general as a process of grouping objects which resemble one another in terms of their properties. But this statement, though true, is far too vague to be useful: we have to give it a cash value. However, we can more usefully, and somewhat more specifically, say that classification is both an information-losing process and an information-gaining process. That is to say, we start with empirical facts about objects and properties, for example, that object *a* has properties 1, 2, and 4, that object *b* has properties 2, 3, 4, and 5, that *a* shares properties 1 and 2 with *c* and 2 and 4 with *d*, that *b* shares 2 and 3 with *c* and 1, 2, and 3 with *d*, and so on. What we want to do is trade all this detailed information for a more general statement to the effect that these objects are all alike, that is that *a*, *b*, *c*, and *d* are all members of a class, because they share properties, irrespective of exactly how they share them. We thus lose information, or rather throw it away, because we forget all about the particular property relationships among the individual objects. But we gain because we make explicit the fact that some objects are alike. Of course this fact is contained in the original data; but the object of classification is to bring it out when it may be far from obvious. We can, however, achieve an information gain of a more important kind in return for the loss of the original detailed information. This is that any member of a class can be treated as if it possessed certain properties – the class-characterizing ones – even if we did not know whether it did to start with. The fact that an object is a member of a class enables us to make an inference about it.

This can be illustrated by the thesaurus case. To start with, we merely know that certain words occur in certain documents. We then say that some words are alike, in being members of the same class, because they tend to co-occur in the same documents; and we can finally, in treating the members of the class as substitutible for one another, behave as if all the words concerned occurred in all the documents in which any member of the class occurs. Grouping the words means that we assume that they all could perfectly well have occurred in any of the documents on which the class is based.

Now this argument says a little more about what a classification is, but unfortunately the modes of treating data which can be described as classificatory are still very varied. The difficulty is precisely to relate them to one another and to characterize classification in general in a useful and non-vacuous manner. Nevertheless, we can say that some sorts of activity do not produce a classification in the sense we are trying to explicate. One such area of activity is that represented by scaling techniques like the Shepherd/Kruskal multi-dimensional scaling procedure (Kruskal, 1964a, b). Scaling techniques are designed to simplify data like classification methods, but they are not divisive. We can say that classification makes those divisions that scaling is anxious to avoid. With a scaling we cannot say these things are all alike, and different from those things. A second apparently comparable area is that represented by factor analysis (Borko, 1965). Here we are nearer to classification than we are with scaling, but it can be said that a classification is implicit in the results of factor analysis rather than explicit. If we say that scaling is intended to present data in a form in which we can understand it, while classification is intended to understand data, we can say that factor analysis provides us with good properties, that is ones which are valuable in forming a classification. In both scaling and factor analysis we throw some information away, but we do not gain as much as we do with a classification in return.

But this is, as it were, a negative attempt to characterize classification. If we try to be more positive about it, it appears that we can categorize classifications according to certain general properties, or characters (reserving the word "property" for a feature of an object). These characters can be defined in terms of answers to three questions which can be asked about all classifications, as follows. The first question asks what the relation between the properties of objects and classes is. The answer is that we can have classes which are either monothetic or polythetic. If a class is monothetic this means that all its members possess the same common property or properties, which is not true of polythetic classes. The second question concerns the relation between objects and classes, where the answer is that we can have classes which are exclusive or overlapping. If objects are assigned to one class only we have exclusive classes, while if they are assigned to more than one we have overlapping classes. The third question concerns the relation between classes, which gives us ordered or unordered classifications. If a classification is ordered this means that the classes in it are systematically related to one another, which does not hold for unordered classifications.

We therefore obtain a pair of alternatives under three headings which can be listed as follows:

(1) *Relation between properties and classes*:
- monothetic; and
- polythetic.

(2) *Relation between objects and classes*:
- exclusive; and
- overlapping.

(3) *Relation between classes and classes*:
- ordered; and
- unordered.

As far as I can see, these six characters exhaust the possibilities as far as the overall features of classifications go: a specific classification must represent a choice of one of the alternatives under each head, each pair being mutually exclusive. (Notice that this remark applies to the principles on which grouping is based, and not to any actual groups: a procedure designed to allow overlapping classes may in fact generate exclusive ones for a given set of objects.) At the same time, all the possible combinations of choices can generate classifications. We can, that is, say that classifications can be divided into the eight sorts corresponding to the permissible combinations of characters as shown in Figure 1.

Clearly, one would like to be able to say that anything worthy of the name classification is covered by this range, and that no treatment of a set of objects it permits is not a classification. However, this is not easy to demonstrate, and the whole categorization rests on the prior notions of object, property and, in particular, class, which we have not defined but assumed. A rather different point is that even if we know what objects, properties, and classes are, and are in a position to demonstrate that classifications are covered by these categories, this would not be very informative because we are still talking at too high a level of generality. Each of the eight sorts of classification covers a whole range of actual classification methods or class definitions, and it is on these that we have to focus our attention. We can indeed distinguish three levels of reference in talking about classification. We can talk, broadly, about classifications of different sorts according to the scheme just presented. However, to obtain a classification of a particular sort we need what may be called either a classification method or a class definition, which specifies the precise base on which objects are to be grouped. Finally, we need an algorithm which enables us to find classes on a given definition for a particular body of data. The important point here is that different methods or definitions generate classifications of the same sort, and further, that different algorithms can be used to generate classifications based on the same definition.

	1		2		3	
	a	*b*	*c*	*d*	*e*	*f*
I	x		x		x	
II	x		x			x
III	x			x	x	
IV	x			x		x
V		x	x		x	
VI		x	x			x
VII		x		x	x	
VIII		x		x		x

Figure 1.

These distinctions can be illustrated by an example. For instance, suppose that we want classes of the polythetic, multiple, unordered sort. Given this initial interest, we may then define a class either as a set of objects which are all connected with one another, in the sense that each pair shares at least one property; or we may define a class as a set having more internal connections between members than external connections between members and non-members, where a connection again represents one or more shared properties for the pair of objects in question. Finally, if we opt for the latter, we may look for classes either by following a procedure which adds objects to a set started from one object, or by adopting a procedure which adjusts an initial random partition of the universe of objects until the objects on a specified side satisfy our requirement.

At this point it is necessary to emphasize the fact that whether we construct one sort of classification or another for a given set of objects has nothing to do with the nature of the data in question. There is no one correct or natural way of classifying a universe of objects. This has to be emphasized because in many cases remarks are made about the relative merits of classifications as if classification reflects something intrinsic to a set of objects, where it in fact reflects the frame of reference of the person seeking a classification. We can perfectly reasonably say that we can in principle apply any method of classification to, and hence obtain a classification of any sort for, a given body of data. Thus in Figure 2 in which the similarity between objects is represented by two-dimensional proximity, the dotted and solid lines represent two equally plausible classifications which differ because they are based on different views of what a classification should be like. The dotted classification follows from the view that classes should be exclusive, while the solid one is a natural consequence of allowing overlap among classes. The alternatives are, however, only associated with different interests on the part of the classifiers: there is nothing in the objects themselves to justify the choice of one classification rather than the other.

The fact that it is legitimate to classify the same set of objects in different ways is particularly important when we come to consider the question of evaluation. When we have constructed a classification we naturally ask whether it is a good one; indeed in practice we often construct several alternative classifications, for a given body of data, so that we ask which is the best of these. What, then, does asking and answering these questions involve? How do we reconcile the assertion that we can perfectly well classify the same set of objects in different ways with the statement that one result is best?

Essentially we can evaluate classifications on two different bases. We may on the one hand have a complete formal theory of classification, buttressed by a formal theory of data, so that for a given set of objects identified as having certain formal

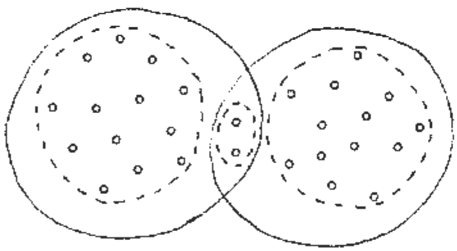

Figure 2.

characteristics, we can conclude that the least violence is done to, or "distortion" is imposed on, the data by such and such a classification method. Or we may on the other hand have a complete formal statement of the purpose for which the classification is required in terms which lead to the choice of the appropriate method. In the first case a distortion measure is associated with the classification theory, so that if we are interested in obtaining a classification purely as a static description of a set of objects, the measure enables us to match a grouping method with our data. The result is then necessarily the best classificatory description of the objects. In the second case we can say that the purpose for which the classification is required itself implies or involves distortion, so that we are then looking for the classification method which minimizes this distortion. The result is then the most appropriate classification for the specified purpose. This may, however, be quite different from the classification we would obtain if we confined ourselves to a description.

It is essential, in other words, to distinguish classification as an abstract process, where internal criteria only are to be satisfied, from classification as response to an objective, where external requirements have to be satisfied. In both cases, however, we can only be sure of our results if we have the formal apparatus for expressing the whole; and the problem is just that of formulating abstract classificatory requirements in the first place, or of reformulating purpose requirements in classificatory terms in the second (assuming we can identify our purpose requirements, which is of course itself a problem). We should indeed say that our purpose requirements should be expressed in abstract classificatory terms in the second case. For while it is in principle true that the requirements of a given purpose can override the formal requirements of a classificatory theory, it is most unlikely that our purpose requirements would be properly satisfied by a classificatory method which did not also satisfy the abstract requirements associated with a descriptive classification theory. This is of course an absolute remark: it must be allowed that useful results can be obtained in practice with theoretically quite defective grouping procedures. It is simply that we can certainly do better as far as our purposes are concerned with methods which are also theoretically satisfactory.

Unfortunately we do not have a complete formal data and classification theory, and equally in many cases have no sufficiently precise statement of our purpose. Many people have dreamed up specific grouping methods, particularly for producing exclusive, ordered classifications for biological material, but very little has been done on the problem of evaluating these methods and relating them to one another within the framework of an overall classification theory in a rigorous and productive way. For a theory of classification worthy of the name we require at least a list of the criteria which must be satisfied by any theory of classification if it is to be formally adequate, which can then be used to sort out the various methods which have been proposed. But it is difficult to provide suitable criteria, and it may also be difficult to show that a given method satisfies them. A truly effective classification theory should, however, go further and show how different classification methods are related to one another, so that the consequences of choosing one method rather than another in different circumstances can be predicted.

Clearly this is a very tall order, particularly when we remember that the classification theory should include a formal data theory, and also a treatment of similarity or dissimilarity measures, since these may be applied to the initial object-property information to provide the actual statements about the relations between pairs of objects which are input to the grouping procedure. Some work has been done on these problems, but chiefly in connection with methods for producing ordered classifications which, it can be said, represent a comparatively tractable sub-area. Thus Jardine and Sibson (1968) have proposed a set of criteria for a classification theory, and have considered ordered grouping techniques in relation to these. A fair amount is also known about similarity and dissimilarity coefficients and their relatives. But for those interested in unordered definitions, as we are in retrieval, the current position is very unsatisfactory. It is not difficult to see that methods which have been proposed may well be theoretically unsatisfactory – for instance clump definitions depend on the dubious notion of adding similarities; but there are no obvious lines to follow which may lead to better ones.

I am at any rate unable to put forward anything more acceptable here. So it seems to me that the best approach to the problem of trying to establish satisfactory methods is to look a little further at what we want a classification theory to do, and then again at our retrieval purpose.

For example, if we consider the question of adequacy criteria, what sort of requirements should we impose on a classification procedure? Jardine and Sibson's list – somewhat improperly expressed informally, and generalized – is as follows:

(1) that a unique result should be obtained for a given body of data;
(2) that if the data are already classified, this classification should be preserved;
(3) that the procedure should be independent of the way the objects are named;
(4) that the method is independent of scale;
(5) that the method minimizes distortion;
(6) that the results should not be radically affected by small changes in the data; and
(7) that maximally-connected sets of objects should not be split.

Requirements like these look very sensible, though it should be noticed, for example, that they imply that we have already supplied a satisfactory distortion measure. In addition, we have to consider underlying requirements, immediately about similarity or dissimilarity coefficients, and ultimately about data. Thus we may impose on a similarity coefficient the requirements:

(1) that identical objects have maximum similarity; and
(2) that objects with complementary property distributions should have minimum similarity.

With regard to the basic data we may wish to formulate requirements which must be satisfied before they are acceptable as input to the classification process, for example:

(1) that the property assignments are unambiguous; and
(2) that the properties are independent of one another.

In addition, we need a companion to the classificatory distortion measure which is associated with the similarity computation and grouping process. This is a data classifiability measure. Our distortion measure tells us how much classes distort the data, but we would also like to know whether an attempt to find groups would be a mistake because there are none to be found, for instance if we have random data or a set of disjoint objects.

But though all these requirements have to be satisfied by a grouping procedure as a whole, it does not follow that any process which satisfies them is a classification procedure: they are necessary but not sufficient conditions for a classification. These requirements ensure that any classes found will be proper ones, but we have also to ensure that we find classes. What, in other words, makes a data-processing technique a classification method and not, for example, a sorting procedure? Here we come to the root problem of explicating the notion of classification, and it is far from easy to say anything helpful about it. We can say that classification involves three distinct ideas: that we should divide the universe of objects; that we should do this in such a way that the subsets into which the objects fall are held together by likeness among their members, and that the resulting description of the objects in terms of their class membership should be simpler than their original description in terms of properties. But it is extremely difficult to make these statements more precise: thus we may in general feel, for instance, that a classification of ten objects which consists of ten classes each consisting of one of the one-member subsets of the universe is not a classification in the proper sense, and more generally that if there are as many classes as objects, this is unsatisfactory. But how to formulate a simplicity or separation criterion such that it would not inhibit appropriate classifications of some bodies of data, is another question. There could, after all, be a set of ten objects which was properly represented by the classification just mentioned. Notice too that though we may wish to define grouping procedures in such a manner that they generate classifications of one or other of the sorts mentioned earlier, saying that we want a classification with such and such of the broad characters in question will not necessarily give us methods which satisfy the general classificatory criteria just discussed.

The depression induced by the lack of a substantive classification theory, and even more by the suspicion that a sufficiently comprehensive theory would be virtually content-free, is not dispelled when we turn to the problems which arise when we think about classification for a purpose, and specifically classification for retrieval purposes. First we have to state our purpose precisely; and then we have to translate it into classificatory terms. But some purposes are not amenable to precise statements, for example that of the research worker who wants a classification to help him think about some material; and some purposes, though they can be stated more precisely, may still be very difficult to translate. Thus we may, for instance, want a classification of books by their sizes which enables us to use shelf space economically while preserving subject groupings to a sufficient extent: this is a more definite requirement than the previous one, but is still difficult to put in formal classificatory terms.

In this connection, the retrieval application is of particular interest because it produces problems at all levels. Thus to start with, what we are looking for are classes of retrieval-optimizing keywords, i.e. sets of terms which are substitutible because they preferentially promote the retrieval of relevant documents. But relevance is a subjective notion: relevance judgements are made by individual users. We may, however, observe that users behave with statistical consistency. But it is still true that relevance has to do with the content or message of a document, and this is inaccessible to us. We have, as we saw earlier, to base our classes of substitutible terms on document-occurrence information because this is all we have available, instead of on unobtainable facts about what may be called the relevance properties of words. We are thus in the position that we *cannot* state our purpose requirements directly in a form which can be exploited in the construction of our classification. We can only state it indirectly, by saying we want classes of text-related words, and make the assumption that these classes will in fact give us the results we would have obtained if we had used relevance information. But the whole process is very unsatisfactory because if our classification is found wanting in terms of retrieval performance, it is not obvious how we should revise it. However, a further difficulty arises in that the mere statement that we want classes of substitutible terms based on document texts is not a very good guide to the choice of a grouping procedure. It is not precise enough. It does not tell us whether we want tightly-connected classes or loosely-connected ones, for example. We are in an awkward position, in that even if we have a well-organized set of classification methods to which we can refer our admittedly surrogate purpose requirement, we do not have a sufficiently specific requirement. So that when we take into account the fact that the well-organized set of classification methods is missing, and that all we have is a congerie of ill-tried suggestions, it will be clear that embarking on automatic classification for retrieval purposes is very much like marching into a bog in the dark with no boots on.

What, then, should our method of working be, given that the need for automatic retrieval classifications remains, and indeed becomes more pressing? The only possibility seems to be the compromise one. We should on the one hand seek to encourage the development of a classification theory, at least for the sub-area represented by the sort of classification we are particularly interested in; and at the same time, we should carry out as many experiments as we can, in as systematic a manner as possible, to try to clarify our needs. These will necessarily depend on current, unsatisfactory grouping procedures; but if the known defects of these procedures are borne in mind, and they are used within the framework of controlled and systematic sets of experiments, we should be able to make some valuable inferences about what classifications for retrieval purposes should be like, and also, hopefully, to get some ideas about better classification procedures.

References

Borko, H. (1965), "Research in computer-based classification systems", in Atherton, P. (Ed.), *Classification Research*, Munksgaard, Copenhagen, pp. 220-57.

Jardine, N. and Sibson, R. (1968), "A model for taxonomy", *Mathematical Biosciences*, Vol. 2, pp. 465-82.

Kruskal, J.B. (1964a), "Multidimensional scaling by optimising goodness of fit to a non-metric hypothesis", *Psychometrika*, Vol. 29, pp. 1-27.

Kruskal, J.B. (1964b), "Non-metric multidimensional scaling: a numerical method", *Psychometrika*, Vol. 29, pp. 115-29.

Needham, R.M. (1963), "A method for using computers in information classification", in Popplewell, C. (Ed.), *Information Processing*, 62, North-Holland, Amsterdam, pp. 284-7.

Spärck Jones, K. (n.d.), "The theory of clumps", *Encyclopedia of Library and Information Science* (in press).

Further reading

Spärck Jones, K. (n.d.), *Automatic Keyword Classification and Information Retrieval*, Butterworths, London (in press).