



## 60 YEARS OF THE BEST IN INFORMATION RESEARCH

# A substantive theory of classification for information retrieval

Birger Hjørland and Karsten Nissen Pedersen

*Royal School of Library and Information Science, Copenhagen, Denmark*

### Abstract

**Purpose** – To suggest that a theory of classification for information retrieval (IR), asked for by Spärck Jones in a 1970 paper, presupposes a full implementation of a pragmatic understanding. Part of the *Journal of Documentation* celebration, “60 years of the best in information research”.

**Design/methodology/approach** – Literature-based conceptual analysis, taking Spärck Jones as its starting-point. Analysis involves distinctions between “positivism” and “pragmatism” and “classical” versus Kuhnian understandings of concepts.

**Findings** – Classification, both manual and automatic, for retrieval benefits from drawing upon a combination of qualitative and quantitative techniques, a consideration of theories of meaning, and the adding of top-down approaches to IR in which divisions of labour, domains, traditions, genres, document architectures etc. are included as analytical elements and in which specific IR algorithms are based on the examination of specific literatures. Introduces an example illustrating the consequences of a full implementation of a pragmatist understanding when handling homonyms.

**Practical implications** – Outlines how to classify from a pragmatic-philosophical point of view.

**Originality/value** – Provides, emphasizing a pragmatic understanding, insights of importance to classification for retrieval, both manual and automatic.

**Keywords** Classification, Semantics, Information retrieval

**Paper type** Conceptual paper

### Introduction

Spärck Jones (1970) pointed out that for information retrieval (IR), theories of classification are inadequate and have not been sufficiently considered. She pointed out that a substantive theory of classification is needed but does not exist. Although many different approaches have been tried, this may still be the case in 2005. Spärck Jones' own thoughts can be interpreted as oscillating between a “positivist” and a “pragmatic” understanding, and this paper tries to demonstrate that a full implementation of the pragmatic view may suggest how to construct the theory needed. Spärck Jones' paper focuses on automatic methods for classification, but the issues addressed seem fundamental for any general theory of classification, intellectual or automated. Inspired by Spärck Jones (1970), this paper outlines the authors' vision of classification theory.

“Some thoughts on classification for retrieval” by Spärck Jones (1970) was first published in *Journal of Documentation*, Vol. 26, No. 2, pp. 89-101. The paper has been republished in this issue as part of a series of articles celebrating 60 years of the best in information research in *Journal of Documentation*.



---

### An overall description of Spärck Jones (1970) and its reception

When renowned information scientist Karen Spärck Jones won the 2002 ASIST Award of Merit, part of the motivation was that she "... has made outstanding theoretical contributions to IR and natural language processing and has built upon this theoretical framework through numerous experiments" (Spärck Jones, 2003, p. 12). Recently, Robertson (2004) acknowledged one of these contributions, namely what has subsequently become known as the inverse document frequency (Spärck Jones, 1972), which has been widely recognized, e.g. by developers of web search engines. Besides, she has been instrumental in the design and implementation of Text REtrieval Conference (TREC) and has demonstrated a lifelong preoccupation with and passion for IR and natural language processing experiments (Spärck Jones, 2003, p. 12).

Spärck Jones (1970) published a paper, "Some thoughts on classification for retrieval", in *Journal of Documentation*. The paper displays Spärck Jones' interests in both theoretical and experimental aspects of classification, given her task of coming up with an automatically constructed classification for retrieval. Spärck Jones thus oscillates between considerations concerning theory and experimentation. Subsequent to the 1970 paper, Spärck Jones has emphasised the experimental aspect in her research, but we would like to draw attention to the theoretical aspect of the 1970 paper because we find that Spärck Jones, at a very early time, touched upon issues which are as important and unresolved today as they were back then. Reflecting the theoretical and experimental aspects of classification, Spärck Jones discusses two interrelated questions, namely: what sort of classification (or classification method) to look for when in want of an automatically constructed one? And how to exploit classification theory given the overall purpose of retrieving documents relevant to requests? Spärck Jones (1970, p. 90) does not seek to come up with answers to these questions, but suggests that discussing these questions may help "...organizing experimental research in this area".

As her starting point, Spärck Jones introduces an eight-wise categorization of classifications based on three pairs of mutually exclusive pairs of characteristics: first, a monothetic classification is one where all members of each class share one or more properties, while a polythetic classification is one where members of a class do not necessarily share one or more common properties. Second, an overlapping classification is one in which objects may appear in more than one class, while, in contrast, an exclusive classification only allows objects to appear in one class. Third, an ordered classification is one in which there are some systematic relationships between classes, whereas an unordered classification presupposes no such systematic relationships between classes. Spärck Jones takes these  $2 \times 2 \times 2 = 8$  categories of classifications to exhaust the possibilities, and subsequently van Rijsbergen (1979, p. 28) has described the categorisation as "...a very clear intuitive break down of classification methods in terms of some general characteristics of the resulting classificatory system".

Spärck Jones (1970, pp. 91-2) argues that, given her purpose of automatic construction of classifications for retrieval purposes, she is in want of a polythetic, overlapping, and unordered classification. However, and within the confines of this category, there is room for numerous classifications or classification methods given a set of objects because "... there is nothing in the objects themselves to justify the choice of one classification rather than the other" (Spärck Jones, 1970, p. 96). At the

same time, the overall retrieval purpose suggests some classification to be the best one and Spärck Jones (1970, p. 97) therefore asks: “How to 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?” This leads Spärck Jones (1970, p. 97) to observe that essentially classifications can be evaluated on two different bases: either on basis of a “complete theory of classification, buttressed by a formal theory of data” or on basis of a “complete formal statement of the purpose for which the classification is required”.

As Spärck Jones does not attempt to come up with answers to her initial questions, this distinction between classification as an abstract process and as a response to an objective leads Spärck Jones (1970, p. 101) to suggest a way – or rather two ways – of working, based on the assumption that the need for automatic retrieval remains. Spärck Jones (1970, p. 101) suggests compromising a way of working:

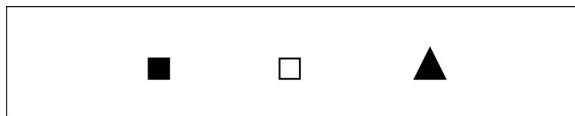
We should on the one hand seek to encourage the development of a classification theory, at least for the subarea represented by the sort of classification we are particularly interested in; and at the same time, we should carry out as many experiments as we can, in as systematic a manner as possible, to try to clarify our needs.

We do not believe that anybody can disagree with this compromise, but we do believe that a stronger emphasis on the pragmatic aspect of classification is required in order to provide the necessary basis also for further experimentation[1].

Inherent in the distinction between classification as an abstract (or empirical) process and classification as being in response to an objective, we recognise a related distinction between positivism and pragmatism. We take classification viewed as an abstract or empirical process to indicate a positivist understanding, while classification in response to an objective signifies a pragmatic understanding. We thus suggest it to be helpful to consider the Spärck Jones’ questions from positivist and pragmatist perspectives respectively.

### Positivism versus pragmatism

Spärck Jones (1970, p. 89) observes “. . . that 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”. This observation is in our view a clear indication of a pragmatic understanding. It is rather simple to demonstrate why it must be true. Below, three figures, namely two squares and a triangle, are presented (Figure 1). There are also two black figures and a white one. The three figures may be classified according either to form or to colour. There is no natural or best way to decide whether form or colour is the most important property to apply when classifying the figures; whether squares should form a class while triangles are excluded or whether black figures should form a class while white figures are excluded. It simply depends on the purpose of the classification.



**Figure 1.**  
Criteria for classification

---

---

We accordingly suggest that a classification is always required for a purpose, why a consideration of that purpose is the most important part of the methodology of information science (IS). Whether or not that purpose can be stated formally is quite another question. We have to accept working at the level of preciseness that we have access to, i.e. this is the condition under which we have to work and we cannot change this by requiring something that is not available to us. This does not, however, change the fundamental insight: we have to base our classifications on knowledge about their purposes. The method of investigating such purposes may be termed pragmatic analysis (or pragmatic method). By implication, the pragmatic method is the fundamental method in classification research (and probably in all of IS as well). The pragmatic method emphasizes the investigation of goals, purposes, interests, and values (Hjørland, 2002, p. 269). We believe that many researchers hesitate to fully accept and implement this pragmatic insight because they either are overwhelmed by the enormity of the task or would like to appear as neutral experts. Our response to such an argument is first that the cost may be that we do not use the most appropriate methods, and second that neither our theoretical nor our empirical knowledge can advance properly if we do not dare to approach the problem in an appropriate way.

In spite of the important acknowledgement of the need of a pragmatic approach, Spärck Jones (1970) seems to favour a view that may be in conflict with the pragmatic way of understanding. In many ways, the dominant view in her paper may be termed “positivist”[2]. By positivist we mean the ideal that scientists should only be interested in “facts”, that they should keep to pure observations, logical deductions, and formal models, while ignoring issues related to interpretation and meaning as well as goals, purposes, and values. The most important distinction between positivism and non-positivism (including pragmatism) may be related to the aforementioned distinction between theory and observation. The most important critique of positivism has pointed out that our observations are theory-dependent. This critique has been raised by many different scientists and philosophers, including Albert Einstein, Karl Popper, Thomas Kuhn, hermeneutical philosophers like H.G. Gadamer, and by the founders of pragmatic philosophy (Charles Sanders Peirce, William James and John Dewey). This critique is thus rather well known and well accepted today, but in actual IS research this understanding is often not implemented and has mostly been ignored.

Classifications of any kind of objects are based on the properties of those objects. This is the point of departure for Spärck Jones (1970) as well as for us. An important point is, however, that the properties are only available to us on the basis of some descriptions (this distinction is emphasized by van Rijsbergen (1979), although he does not draw the implications that we do in this paper).

Spärck Jones’ (1970, p. 93) statement, “we start with empirical facts about objects and properties”, is just one among many signs that her paper may be viewed as being grounded in positivist norms of science. There is no indication that our knowledge of the objects and their properties comes from descriptions and that those descriptions may vary from one observer to another or that they may be theory-dependent. If we accept that our knowledge about objects and properties comes from their descriptions, e.g. in the literature, and that they may be more or less dependent on the point of view of some observer, e.g. an author, then we have a very different perspective. We may

---

still subscribe to the positivist ideal of inter-subjectivity, but the costs may be high: if observations are theory-dependent, only very trivial observations may be shared among all observers, and consequently we have to base our classifications on trivial descriptions rather than on important or essential descriptions.

A further complication is that the descriptions may exhibit various degrees of synonymy and homonymy; that is, the same objects and properties may be labelled differently and different objects and properties may be labelled similarly. As is recognized in semantics, pure synonyms and homonyms are rare; rather terms may be regarded as more or less synonymous or homonymous (or otherwise related) depending on perspectives (Frohmann, 1983). Hence, this property of the descriptions may also be theory-laden.

We will not discuss the possibility of basing classifications on trivial descriptions, and the consequence is that we have to determine criteria of how to select descriptions and how to use them. We are far from the situation where objects and properties are just “given”. Different human interests emphasize different properties of objects: pharmacology and chemistry, for example, emphasize different properties of the same chemical elements (a chemical database emphasizes structural descriptions while a pharmacological database emphasizes medical effects). Furthermore, and within all fields, different theories and “paradigms” also emphasize different properties. A typical database, on which IR-experiments are performed, should be conceived as a merging of different descriptions serving different purposes. The average point of view is not optimal to any purpose, why statistical methods based on average samples of documents and users are probably not very helpful.

The consequences of the pragmatic insight are also that any classification or IR-system is more or less suited in relation to given goals, purposes, and values. Web search engines, for example, are not neutral tools, which are just to be evaluated on the basis of technological efficiency; rather they should be seen as supportive of some interests at the expense of other interests (Introna and Nissenbaum, 2000).

Likewise, Spärck Jones’ eight-wise categorization of classifications is subjectable to the pragmatic principle of being more or less relevant given some purpose. Van Rijsbergen (1979, p. 29) states: “The distinction between monothetic and polythetic is a particularly easy one to make providing the properties are of a simple kind, e.g. binary-state attributes. When the properties are more complex the definitions are rather more difficult to apply, and in any case are rather arbitrary”. For this reason the suggested categorization of classifications may be of limited value to IR in pursuit of automatically constructed classifications for retrieval.

In general, one could say that a specific interest (say that of Scandinavian public libraries) should lead to the design of systems, which are optimal given the interest or purpose and which do not just lead to the acceptance of implicit values inherent in systems that are designed, for example, for commercial purposes. A system designed for a Scandinavian public library should not, for example, tend to identify American commercial websites but should be optimised to identify pages reflecting their cultural and democratic values and purposes.

### **Classification and conceptualization**

Discussing the concept of concepts, Rey (1998, p. 505) points out that “. . . there is considerable disagreement about what exactly a concept is”. Despite such

---

disagreement, he points out that concepts "... seem essential to categorizing the world". Hereby, Rey indicates classifications, concepts, and conceptualizations to be intertwined, and we are to suggest that differing opinions on – or theories of – what is a concept may comply with the previously introduced distinction between positivism and pragmatism.

Andersen *et al.* (1996, pp. 348-51) provide an account of two theories of concepts, namely the "classical" theory (dating back to Aristotle)[3], where concepts are defined by a set of individually necessary and jointly sufficient properties, and Kuhn's (1996, 1970, 1996, p. 349) theory of concepts informed by the so-called "prototype theory" associated with Ludwig Wittgenstein and Eleanor Rosch, where objects in a similarity class "... need bear no more than a family resemblance to their fellows, and hence that the concepts corresponding to these similarity classes are family resemblance concepts"[4]. Kuhn, in rejecting a specific set of properties as the basis for class membership, thus rejects the traditional view that concepts and classes can be defined in terms of necessary and sufficient properties.

According to the traditional view, defining a concept implies being devoted to the task of uncovering plausible candidates for a set of universally accepted necessary and sufficient properties. Andersen *et al.* (1996, p. 348) suggest that the traditional view was predominant in the Western philosophical tradition until the mid-1970s, where Kuhn, inspired by, among others, Wittgenstein, paved the way for a more relativistic understanding. Kuhn (1962, 1970, 1996) has become widely known for his *Structure of scientific revolutions*, but what has become less recognized is that Kuhn, in the wake of the heated debate following *structure*, also suggested a theory of concepts and taxonomies (and thus classification). Developing his theory of concepts, Kuhn was influenced by Wittgenstein's family resemblance account of concepts; which allows objects to instantiate the same concept without, in extreme cases, sharing a single property:

[Wittgenstein] ... clearly envisaged a situation in which although the first instance of a concept shared some features with the second, the second shared different features with the third. In such situations, although it was true that all objects falling under the concept shared some features with their fellows, no single feature or set of features had to be common to all the objects (Andersen *et al.*, 1996, p. 348).

Partly in response to critique[5] of the family resemblance account of concepts, Andersen (2002, p. 99) points out that "... family resemblance concepts form hierarchical structures in which a general concept decomposes into more specific concepts that may again decompose into yet more specific concepts, and so forth – in other words taxonomies"[6]. Hereby, Andersen suggests that classification may be explained systematically from a family resemblance point of view and furthermore (p. 99) argues that the family resemblance account allows for taxonomies being dynamic entities, which may undergo change.

According to Andersen (2002, p. 102) theories or models "... provide the causal and explanatory links that hold individual concepts together and establish taxonomic relations to other concepts". This account differs markedly from the "classical" theory of concepts, where concepts are defined by a set of individually necessary and jointly sufficient properties. Here, that which holds concepts together and that which establishes relations to other concepts are theory-independent.

---

Andersen (2002, pp. 103-8) recounts the Gamow versus Bohr dispute on the nuclear disintegration process as an example of two theories which yield markedly different taxonomies. Thus, taxonomies develop in response to "... a continuous interplay between, on the one hand categorizations and recategorizations of observed phenomena on the basis of characteristics that make them seem similar or dissimilar, and on the other hand the development of models [or theories]" (Andersen, 2002, p. 108).

An important consequence of Kuhn's account of concepts is that although a class is agreed upon, the identification of class and non-class members need not be based on the same characteristics or set of characteristics:

Nothing limits the possible characteristics according to which the objects may be similar or dissimilar. On the contrary, anything one knows about the referents can be used when matching them with terms (Kuhn, 1983). This also means – in contrast to the traditional view – that although different speakers use a given concept to pick out the same similarity class, they need not identify instances and non-instances by the same features (Andersen *et al.*, 1996, p. 350).

Consequently, models or theories are only partial and contextual and no model or theory "... explains all aspects of a taxonomy, instead scientists adopt different models to explain different things. In some cases, these different models employed to solve different questions may turn out to be incompatible or inconsistent" (Andersen, 2002, p. 109). For example, in the Gamow and Bohr case, the theories or models were incompatible.

The agreement between the pragmatic and Kuhn's view on concepts may be summarized this way: according to pragmatism, concepts are ways of classifying the world that serve human purposes. The concept of "food" within a culture serves the purpose of indicating for the people in that culture what is and what is not eatable. Kuhn's theory explains how different observers and different theories or paradigms may classify differently in accordance with the internal logic of the theory or paradigm. If a certain substance X in one culture is regarded food, while not so in another culture, then the semantic relations between X and "food" differ for the two cultures (or theories).

Considering the two differing theories of concepts affirms our previous observation, namely that classifications are theory-dependent and do reflect goals, purposes, and values. Given two concepts A and B it is thus theory-dependent whether their semantic relations are of a given kind or not. In other words: two concepts may be synonyms in one context, but not so in another. Spärck Jones recognizes this last insight, but we believe that Kuhn's theory of concepts and the pragmatic philosophy provide better explanations of why this is the case. It also puts us in a position where we are better suited to predict where, when, and how certain semantic relations may be identified.

### **Synonymy and homonymy[7] in IR (another view on automated classification)**

The problems of synonyms and homonyms are related to the problem of classification because a classification is supposed to unite different terms with the same meaning and to differentiate between different meanings of the same term. Spärck Jones has investigated algorithms that may be used to automatically classify document representations based on the identification of synonyms and homonyms. We here suggest what we believe to be a new approach to this problem. An approach that shares

---

some characteristics with the use of test collections (Spärck Jones and van Rijsbergen, 1976) but which is more interpretative and in a way puts things upside down[8].

The approach we suggest is based on the study of specific literatures, avoiding any generalization (at least as a starting point). In such a sample of literature, given homographs may be uncovered manually and subsequently it may be examined how different elements of the document representations contribute to differentiate meanings of the same term (see example by Kirkehei, 2004, presented below). The basic idea is that we hereby are able to estimate the value of different discriminators without depending on such things as relevance judgment associated with traditional IR experiments. This makes it possible to construe an algorithm for the solution of the homograph, which, however, is not just one more algorithm among others, but which is the theoretically optimal algorithm given the specific literature and the manually identified homograph. But why prefer such an approach when we do not believe that it is generalizable to other homographs or other literatures? The answer to this question is that we believe that we may gradually learn more and more about specific literatures and about the relatively standardized practices that authors perform (i.e. about genres). In other words, we learn about different ways of expressing oneself in documents and relatively stable patterns[9] which may be exploited in IR. This is quite different from a positivist method of performing one experiment after another, attempting to identify an optimal algorithm for a non-specific purpose or context.

Kirkehei (2004) has examined one homomorphous phrase, “activity theory”, in order to identify which descriptions allow for the best identification of papers reflecting different meanings of this phrase. She made a combined qualitative and quantitative study by examining the phrase “activity theory” in PsycInfo and Social SciSearch. A total of 568 records were found. By examining the records and the literature, she “manually” identified three different senses of this phrase in the two databases. In other words, the phrase represents a homonym covering three meanings. The first important step was to classify the records and to identify the founders and the disciplinary borders of each sense:

- (1) Cultural historical activity theory (341 records. Disciplines: psychology and educational research. Founders: L.S. Wygotsky and A.N. Leontiev).
- (2) Activity theory of aging (47 records. Discipline: gerontology. Founder: R.J. Havighurst).
- (3) Routine activity theory (26 records. Discipline: criminology. Founder: M. Felson).
- (4) Noise and duplicate records.

As argued, one main task for IR is to differentiate between these three meanings of the phrase “activity theory”. The standard way is to make searches and make the users indicate which records are “relevant”. Also inherent in the traditional methodology is the search for some generalized connections between dependent and independent variables.

Kirkehei’s (2004) methodology is different. She starts with a homonym with some manually identified meanings and some important attributes related to the literature of each meaning (its disciplinary borders, its founding authors, its geographical, historical, and linguistic extensions, and so on). The manual classification of the records into groups representing the three senses is supposed to be a highly reliable classification[10]. Then it was examined how good discriminators different elements

---

were. For example: how many records (among the 341) could be identified by examining a set of cited authors (such as L.S. Vygotsky, A.N. Leontev, and Y. Engelstrom)? How many records (of the 479) could be identified by limiting the search to gerontological journals?[11].

Based on such investigations it is possible to design an algorithm which is able to discriminate between the three senses. The point is that this is the best one theoretically possible to construct given the existing literature! Given such an algorithm we know which elements in the literature can be used to discriminate the different senses and we know how much each element is able to contribute and how much any combinations of elements are able to contribute. We have solid knowledge of a nature not provided by mainstream experiments.

Such an algorithm is based on the assumption that different descriptions are more or less suitable when different senses of a homomorphous term are to be identified. This approach is connected to pragmatism because the assumption is that the use of homonyms and genres does not follow universal norms but they are forms of expression developed culturally to adjust to some specific conditions. Because different (sub)cultures develop different genres we do not believe that the same algorithm can be optimal for different literatures or homographs. Because databases typically represent mergers of different descriptions serving different purposes such underlying patterns should be investigated.

We do not assume that other homonyms (e.g. “modulation” or “Aids”) behave the same way. We believe they are used in different discourse communities characterised by different genres and properties, while such other homonyms display other patterns. Many IR researchers would probably criticize this and find the study quite uninteresting from an applied point of view. From a theoretical point of view, we do, however, suggest the opposite critique: that the applied studies may tell whether algorithm A is better than algorithm B given some “average” conditions, but such studies do not accumulate solid knowledge about the underlying structures and mechanisms. From one case study one might go on with similar studies. The point is that the generalization of the theoretically optimal algorithm is an empirical question: one has to examine different literatures and senses. Our understanding of homographs implies that different discourse communities use the same sign without any trouble as long as their literatures are not merged into one database. They use distinct journals and other communication outlets perhaps even without knowing that other discourse communities use the same term with different meanings. When literatures are merged a need to be able to distinguish these merged meanings arises. The job is thus to be able to separate the different meanings. The job for a theory of classification and retrieval is to identify the cues for how best to separate such merged discourses and meanings in order to uncover the relative meaning of different access points in different contexts (Hjørland and Kyllesbech Nielsen, 2001).

What we have suggested above implies that the study of different domains, traditions, genres, and perspectives in literatures might help identify patterns which may be helpful in IR.

### **How to identify goals and criteria for classification**

Among the goals, purposes, and values to be considered by IR systems are commercial, educational, scientific, scholarly, practical, and cultural ones. A public library is, for

---

example, supposed to have a goal that is somewhat different from a typical commercial bookstore. To some degree, one kind of tool (including classifications and algorithms) may satisfy more kinds of needs, but the differences in goals, purposes, and values may also have implications: public libraries and commercial bookstores may disagree on which and how documents are to be described and organized. One might imagine, for example, that the booksellers' descriptions and classifications to a higher degree would emphasize "popular" aspects of a book, while the public library to a higher degree would reflect some cultural values (which may perhaps ideally be influenced by cultural studies taught in schools of library and IS).

If the goals come out clear, it is easier to describe what documents may further the goals than if the goals are unclear, vague, or, as is often the case, conflicting. At the individual level, persons may pursue quite idiosyncratic goals, but such unclear or individual goals cannot serve as a sound basis for classification. There have to be either a larger group of people sharing identical goals or there have to be some ideal goals that can be used as basis for classification.

Commonly, "user studies" in IS are supposed to offer an appropriate way of uncovering users' needs and thus a basis for classification. Such studies face their own problems because users are normally not able to specify criteria in a literature of which they are not very knowledgeable (if they were, there would probably be little reason looking or searching for it). Another basis for classification is the principle of literary warrant (Hulme, 1911), but unfortunately, research has not properly addressed how to interpret the literature in order to identify or determine criteria on which classification may be based. Often it has been assumed that subject specialists can just "read off" the needed information from the literature (which is a kind of positivist understanding). Not much research has examined the basis on which they interpret the literature.

We suggest that different theories, views, and paradigms form scientific and cultural discourses, which are more or less inherent in the literature to be classified. Such theories, views and paradigms may be more or less vague, more or less conflicting, and more or less explicit in the literature. Kinds of philosophical analysis and "science studies" may uncover the different theories and positions in the literature and make them clear enough to be useful as criteria for classification and IR[12].

Turning to medicine as an example[13], it is common to think of this field as one in which there is consensus concerning basic criteria for claims (such as, for example, double blind clinical trials). To the extent such a consensus exists it is possible to deduce criteria for clinical practice and also for the selection of clinical information (and thus for relevance). Medical examples are well suited for illuminating goals and consequences of IR and classification because wrong information may have fatal consequences, insofar as patients are not cured or may even die. The overall goal of all medical research is thus to help cure patients and this goal is inherent in the medical literature and does not differ from the goal of medical libraries. Databases and classifications are consequently to assist in curing patients by identifying and providing appropriate literature. There may be different theories, views, or paradigms that display different opinions on how a cure is best accomplished. Such different views tend to be more or less represented in the literature (especially, the more theoretical and philosophical parts of the literature) and they have different implications for how to classify for retrieval.

To the extent that explicit criteria can be formulated, drugs, as well as documents on drugs, may be classified according to their medical effects (e.g. antidepressiva,

---

anti-inflammatory drugs, blood pressure drugs, etc.). However, if there is no consensus on whether a given drug is efficient given a disease, it becomes complicated. In other words, complications appear if different views or theories exist on whether a given drug (or literature on this drug) is or is not to be classified as a cure to a disease.

The two most extreme theories may be labelled “established” versus “alternative” medicine respectively. Within “established medicine” different interpretations of the roles of experimentally based or clinically based claims may exist and different experiments and experiences may be interpreted somewhat differently by different groups of people (and in different specialities within health science). The criteria which medicine applies when determining whether a cure or drug or a piece of information is relevant are inherent in the way medical education, research, and practice is carried out and tend to be more or less explicitly considered in the literature on the methodology and epistemology of medicine. If different views are identifiable in this literature, then classification as well as IR face conflicting bases on which systems may be founded.

In opposition to a positivist understanding, we believe that such conflicting views will always be present to some extent, especially when databases are merged[14]. The conflict between established and alternative medicine is perhaps not too much of a problem[15], but within “established” medicine different theories and paradigms may also be identified. They may sometimes be difficult to grasp and identify, but our point, as indicated earlier on, is that they are probably the best or even the only criteria we have on which to base classifications. For example, a movement within medicine known as “evidence based medicine” (EBM) has gained much influence. That the MEDLINE database is being re-indexed based on this philosophy[16] is, in our opinion, a strong indication of the role of paradigms in the description and classification of documents. The criteria for how to search and how to index developed in relation to EBM have probably led to the formulation of the most explicit criteria of relevance and of classification existing in this field[17]. If alternatives to EBM can be formulated, such views will provide equally alternative criteria of description and classification.

The medicine example illustrates partly that words, concepts, documents, genres, etc. come about as the result of an ongoing (historical) process centred on an ultimate goal, namely curing the patient. Different opinions on how this goal is furthered are reflected in the medical literature and tend to express themselves at all levels: in the use of words, in the meaning of concepts, in what is considered synonyms or homographs, in relevance criteria, in the forms of publications and communication, etc. Any classification intended for IR has to be more or less in accordance or conflict with views inherent in the literature and the development of such a classification cannot ignore such views. On the contrary, it has to be based on an understanding and a negotiation between them. The goal to be satisfied by a classification system for IR in medicine cannot be based on its own context-independent criteria for classification but has to subordinate itself to goals inherent in the knowledge production in the field. The meanings of terms and the relations between terms are what classification is basically about and such meanings and relations are negotiated in the broader knowledge producing domains and discourse communities[18].

### **Summary of our theoretical view on classification**

Classification is the sorting of objects based on some criteria selected among the properties of the classified objects. The basic quality of a classification is the basis on which the criteria have been chosen, motivated, and substantiated.

---

The way we conceptualise an object depends on our pre-understanding, of our social-cultural background, of our domain-specific knowledge, and of our theoretical outlook. People with different theoretical outlooks tend to describe and classify objects differently. Library and information science (LIS) is not primarily involved with descriptive studies of how people actually classify things. This is the focus of, among other fields, psychology and social anthropology. LIS is primarily concerned with a normative theory of classification: which criteria should be used to classify documents in order to optimise IR?

LIS classification is thus tied to some objectives. Those objectives are generally not specific to LIS, but form part of the goals inherent in the knowledge producing and knowledge utilizing systems and processes in society. Medical research, for example, is aimed at improving health in society. Inherent in medical research are specific theories, assumptions, and views, which influence the way we describe and classify parts of the world (e.g. illnesses, and drugs). Before documents have to be classified in LIS environments, meanings are developed and built into concepts, terminology, forms of publication, and the whole scientific communication system (partly also into the public sphere). When documents are classified in a LIS-context, the basis for conceptualising the subjects of the documents is given in the meanings available in the domain or in the broader cultural environment. Our point is, in other words, that LIS classification should not be based on a trivial or naïve description of the documents, but on the contrary, it should be built on knowledge on the broader meaning-producing contexts.

Knowledge is always produced from specific points of view and tends to emphasize certain interests at the expense of other interests. When classifying documents information specialists are confronting a literature containing more or less different conceptions or paradigms, each of which emphasizes different aspects and goals. This is visible, for example, in different tendencies to cite other documents (Hjørland, 2002). A library may be devoted to a specific view (e.g. a feminist view) but in all cases it is important that users of a classification have access to different views and have access to information not only from a dominant ideology but also from major minority views. It is not possible to be neutral, but is absolutely unacceptable to hide different views and to suppress the users' ability to develop their own points of view. There is thus a deep and unavoidable dilemma connected to classification for IR: the theoretical and ideological basis on which is should be built (a "neutral" system tends to be based on a dominant view and is usually not seen as neutral from outside perspectives[19]).

Our best answer to this dilemma is that the people developing the classification tools and performing the classifications should be adequately educated and be in a position to differentiate theoretical positions as well as to have a basic understanding of their inherent values and consequences. They should be able to understand different views and to base their classification on a negotiation between different views. Such a view may be termed reflexivity. We will point to a paper by Ørom (2003) as a fine example on how different arts classifications may be subjected to such reflexive evaluation when having the proper theoretical background. LIS personnel, having read and understood this paper (or corresponding papers), should be much better suited to develop classifications for arts (as well as to do the actual classification in any universal classification scheme or in any special scheme for arts) compared to people without such a background.

---

## Conclusion

In her 1970 article, Spärck Jones asked for a theory of classification for IR. She did not consider “manual” theories previously developed in IS, such as, for example, the facet analytic approach. Perhaps, more importantly, she did not consider the developing methods of clustering on basis of co-citations in the area of bibliometrics. This area is theoretically important because the clusters identified are based on how researchers actually cite papers they have found relevant to their research. As indicated in Hjørland (1998, p. 25), researchers’ references may theoretically be viewed as a kind of checklist in relation to the tasks that IR is supposed to carry out. Spärck Jones’ paper belongs to a specific tradition of IR in which she, along with, e.g. van Rijsbergen and Salton, is among the most prominent contributors. Her importance for IR is very big, as documented, for example, by Robertson (2004).

Her claims in the paper are indeed modest. She does not claim to provide any framework for a theory of classification. There may perhaps even be a slight tone of desperation in the text in its search for a theory? We believe that such a theory may well be found, but that people like Spärck Jones with a primary interest in computer applications may be somewhat disappointed when they see it (this may indeed be the reason they cannot find it). This disappointment is not, however, a thing one can find ways to make disappear. There are fundamental conditions involved that we cannot change. For example, Spärck Jones asked for a precise statement of the purpose of a classification. But if people’s purposes are not precise, we cannot change that condition, but we have to work on the premises that we have got.

Our suggestion for a new approach in IR emphasizes a combination of qualitative and quantitative techniques, a consideration of theories of meaning and the adding of top-down approaches to IR in which divisions of labour, domains, traditions, genres, document architectures, and so on are included and in which specific algorithms are based on the examination of specific literatures, not on averages of mixed literatures and users. We claim that a theory of classification is especially connected to science studies and that a fairly good outline of such a theory is published in the present paper.

## Notes

1. Experiments are informed by some kinds of theory and theory is informed by some kind of empirical knowledge. We do not claim that interests in experiments are *per se* more positivist than are interests in theory. See, for example, Hjørland (2005) for more on the meaning of “positivism”.
2. Like many related terms “positivism” may be interpreted differently. Hence the quote marks. For a discussion of our understanding of this term, see Hjørland (2005).
3. Actually, more understandings of the concepts of concept may be referred to as being traditional or classical. Wilson (1997), for example, distinguishes between a traditional (rationalist) doctrine of concepts deriving from Aristotle and a classical empiricist understanding where general concepts are accounted for “. . . without invoking the abstract forms or essences to which the Aristotelians and the rationalists appealed”. We follow the simplified Andersen, Barker and Chen account.
4. As for our account of positivism and pragmatism, we do not do justice to the complexity of the concept of concepts. In this paper we restrict ourselves to following the Andersen, Barker and Chen account.
5. See, e.g. Andersen (2002, pp. 97-8) for a short account, or turn to, e.g. Sutcliffe (1993) for an extensive account including a serious critique of Wittgenstein’s view.

6. Excerpt of note in original: “Not all scientific concepts need form taxonomies in the way indicated here. On the contrary . . .”
7. Lancaster (1986, p. 69) distinguishes between homographs (one string of characters, more meanings), homonyms (one string, more meanings, only one pronunciation) and homophones (more strings of words, more meanings, only one pronunciation). We only deal with writings and according to Lancaster we therefore deal with homographs. Nonetheless we follow the common understanding of what are homonyms (and synonyms). Lancaster (1986, pp. 69-71) further discusses homonymy and points out that fortunately it “. . . is much less of a problem than might appear at first sight”. According to Lancaster, homonymy is mainly a problem at the single-term level and since IR rarely involves single-term search strategies the problem is overrated.
8. In *Automatic Keyword Classification for IR*, Spärck Jones (1971), as well as in numerous other papers, digs further into this problem. A short recent reference to Spärck Jones’ way of handling polysemy in general is Evens (2002, p. 149).
9. Hedlund *et al.* (2001) comprises an example from IR research which is also based on the study of another kind of relatively stable pattern, namely features of the Swedish language. Again, such research provides in our opinion a much more concrete and cumulative kind of knowledge compared to typical IR experiments abstracting from specific domains, languages or genres.
10. Of course any problems in this classification will damage the investigation. The most important problem is to avoid circularity: if the citing of a founder is studied as one among other discriminators, this criterion should not be the criterion used to classify the items at first.
11. The author had no access to full-text papers in electronic versions. In that case it would be obvious to see how good discriminators different co-word combinations would appear to be.
12. This view of conflicting theories and paradigms in the literature is different from the ordinary positivist views that have dominated information science so far, including Spärck Jones (1970) search for “empirical facts”.
13. For another example from arts, see Ørom (2003).
14. As discussed by Hjørland (1998), databases may be considered a merger of different journals while almost all database are, in principle, a merging of different meanings.
15. If “alternative medicine” should be taken seriously (and it is indeed a serious and costly task to built databases and retrieval systems for a field like alternative medicine) then its principles and methods have to be formalized and the whole enterprise tends to become more “established”, perhaps established as a theoretical point of view or as a “paradigm” within medicine.
16. E.g. by adding descriptors like “controlled clinical trial” to documents fulfilling certain requirements (see further: [www.nlm.nih.gov/services/cochrane.html](http://www.nlm.nih.gov/services/cochrane.html)).
17. cf. Hjørland (2001).
18. van Rijsbergen (1986, p. 194) has pointed out that the concept of meaning has been overlooked in IS, while the whole area is in a crisis. The fundamental basis of all the previous work – including his own – is wrong because it has been based on the assumption that a formal notion of meaning is not required to solve the IR problems. We believe this to be an important insight, but we do not believe that either van Rijsbergen himself or the whole IR-community has drawn the same conclusion that we have.
19. It may also be a conservative view that is unsuited to solve the problems of tomorrow, which should be the aim of all research.

---

**References**

- Andersen, H. (2002), "The development of scientific taxonomies", in Magnani, L. and Nersessian, N.J. (Eds), *Model-Based Reasoning: Science, Technology, Values*, Kluwer Academic, New York, NY, pp. 95-111.
- Andersen, H., Barker, P. and Chen, X. (1996), "Kuhn's mature philosophy of science and cognitive psychology", *Philosophical Psychology*, Vol. 9 No. 3, pp. 347-63.
- Evens, M. (2002), "Thesaural relations in information retrieval", in Green, R., Bean, C.A. and Myaeng, S.H. (Eds), *The Semantics of Relationships: An Interdisciplinary Perspective*, Kluwer Academic Publishers, Dordrecht, pp. 143-60.
- Frohmann, B.P. (1983), "An investigation of the semantic bases of some theoretical principles of classification proposed by Austin and the CRG", *Cataloging & Classification Quarterly*, Vol. 4 No. 1, pp. 11-27.
- Hedlund, T., Pirkola, A. and Järvelin, K. (2001), "Aspects of Swedish morphology and semantics from the perspective of mono- and cross-language information retrieval", *Information Processing and Management*, Vol. 37, pp. 147-61.
- Hjørland, B. (1998), "Information retrieval, text composition, and semantics", *Knowledge Organization*, Vol. 25 Nos 1-2, pp. 16-31, available at: [www.db.dk/bh/publikationer/Filer/ir\\_semant\\_2.pdf](http://www.db.dk/bh/publikationer/Filer/ir_semant_2.pdf)
- Hjørland, B. (2001), "Why is meta analysis neglected by information scientists?", *Journal of the American Society for Information Science and Technology*, Vol. 52 No. 13, pp. 1193-4.
- Hjørland, B. (2002), "Epistemology and the socio-cognitive perspective in information science", *Journal of the American Society for Information Science and Technology*, Vol. 53 No. 4, pp. 257-70.
- Hjørland, B. (2005), "Empiricism, rationalism and positivism in library and information science", *Journal of Documentation*, Vol. 61 No. 1, pp. 130-55, available at: [www.db.dk/bh/Empiricism\\_p.130-155%2071705.pdf](http://www.db.dk/bh/Empiricism_p.130-155%2071705.pdf)
- Hjørland, B. and Kylliesbech Nielsen, L. (2001), "Subject access points in electronic retrieval", *Annual Review of Information Science and Technology*, Vol. 35, pp. 249-98.
- Hulme, E.W. (1911), "Principles of book classification: introduction", *Library Association Record*, Vol. 13, pp. 354-8, 389-94, 444-9.
- Introna, L. and Nissenbaum, H. (2000), "Shaping the web: why the politics of search engines matters", *The Information Society*, Vol. 16 No. 3, pp. 1-17, available at: [www.nyu.edu/projects/nissenbaum/papers/searchengines.pdf](http://www.nyu.edu/projects/nissenbaum/papers/searchengines.pdf)
- Kirkehei, I. (2004), "Forslag til nye perspektiver i information retrieval: udvikling af IR-algoritmer med fokus på kvalitative og kulturelle aspekter: Teoretiske og praktiske overvejelser", unpublished Master's thesis (in Danish).
- Kuhn, T.S. (1962, 1970, 1996), *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago, IL.
- Kuhn, T.S. (1983), "Commensurability, comparability, communicability", in Asquith, P.D. and Nickles, T. (Eds), *PSA 1982: Proceedings of the 1982 Biennial Meeting of the Philosophy of Science Association: Symposia, Special Sessions, and Invited Lectures*, Vol. 2, *Philosophy of Science Association, East Lansing, MI*, pp. 669-88.
- Lancaster, F.W. (1986), *Vocabulary Control for Information Retrieval*, 2nd ed., Information Resources Press, Arlington, VA.
- Ørom, A. (2003), "Knowledge organization in the domain of art studies – history, transition and conceptual changes", *Knowledge Organization*, Vol. 30 Nos 3-4, pp. 128-43.

- 
- Rey, G. (1998), "Concepts", in Craig, E. (Ed.), *Routledge Encyclopedia of Philosophy*, Vol. 2, Routledge, London, pp. 505-17.
- Robertson, S. (2004), "Understanding inverse document frequency: on theoretical arguments for IDF", *Journal of Documentation*, Vol. 60 No. 5, pp. 503-20.
- Spärck Jones, K. (1970), "Some thoughts on classification for retrieval", *Journal of Documentation*, Vol. 26 No. 2, pp. 89-101.
- Spärck Jones, K. (1971), *Automatic Keyword Classification for Retrieval*, Butterworths, London.
- Spärck Jones, K. (1972), "A statistical interpretation of term specificity and its application in retrieval", *Journal of Documentation*, Vol. 28 No. 1, pp. 11-21.
- Spärck Jones, K. (2003), "2002 ASIST award of merit: Karen Spärck Jones", *Bulletin of the American Society for Information Science and Technology*, Vol. 29 No. 3, pp. 12-14, available at: [www.asis.org/Bulletin/Feb-03/jones.html](http://www.asis.org/Bulletin/Feb-03/jones.html)
- Spärck Jones, K. and van Rijsbergen, C.J. (1976), "Information retrieval test collections", *Journal of Documentation*, Vol. 32 No. 1, pp. 59-75.
- Sutcliffe, J.P. (1993), "Concept, class, and category in the tradition of Aristotle", in van Mechelen, I., Hampton, J., Michalski, R.S. and Theuns, P. (Eds), *Categories and Concepts: Theoretical Views and Inductive Data Analysis*, Academic Press, London, pp. 35-66.
- van Rijsbergen, C.J. (1979), *Information Retrieval*, 2nd ed., Butterworths, London, available at: [www.dcs.gla.ac.uk/Keith/Chapter.3/Ch.3.html](http://www.dcs.gla.ac.uk/Keith/Chapter.3/Ch.3.html)
- van Rijsbergen, C.J. (1986), "A new theoretical framework for information retrieval, in American society for computing machinery", *Proceedings of the 1986 ACM Conference on Research and Development in Information Retrieval*, ACM Press, New York, NY, pp. 194-200.
- Wilson, F. (1997), "Concepts", in Garrett, D. and Barbanell, E. (Eds), *Encyclopedia of Empiricism*, Fitzroy Dearborn Publishers, London, pp. 73-5.