Journal of Documentation, Vol. 54, No. 5, December 1998

THEORY AND METATHEORY OF INFORMATION SCIENCE: A NEW INTERPRETATION

BIRGER HJØRLAND
bh@db.dk

Royal School of Library and Information Science
Birketinget 6, DK-2300 Copenhagen S, Denmark

This paper analyses the theoretical and the epistemological assumptions of information science (IS). Different views of knowledge underlie all major issues in IS. Epistemological theories have a fundamental impact on theories about users, their cognition and information seeking behaviour, on subject analysis, and on classification. They have also fundamental impact on information retrieval, on the understanding of 'information', on the view of documents and their role in communication, on information selection, on theories about the functions of information systems and on the role of information professionals. IS must be based on epistemological knowledge, which avoids blind alleys and is not outdated. The paper shows limitations in the dominant approaches to IS and proposes alternative viewpoints.

INTRODUCTION

In 1997 Vickery [1] discussed in this journal the metatheory of information science (IS). He found that philosophy is a valid source of new hypotheses, but that many recent expositions remain at too general a level: 'If a metatheory is to prove fruitful, it must make connection with the presuppositions already existing in information science, show their weaknesses and propose alternatives' [1, p. 458]. I certainly agree. In this paper, I shall outline a new theoretical and metatheoretical perspective for IS that fulfils this demand. Vickery's own listing of presuppositions is, however, also very general and it seems not to consider different or conflicting approaches in IS (as, for example, those presented in [2]). From reading his article one can get the impression that facet classification, the cognitive view, bibliometrics, the information retrieval (IR) tradition etc. do not reflect different metatheoretical approaches. I also missed an analysis of the historical context. An historical analysis would be helpful in order to learn from the cumulated experiences. Have there been 'blind alleys'? How can we avoid them in the future? In order to answer these questions, we must consider the basic epistemological assumptions on which IS is based. An analysis should include an interpretation of the development of IS as well as an interpretation of interdisciplinary and philosophical trends of importance to information science.

There are two basic claims in this article. First, that there exists a close relationship between problems and approaches in IS and in epistemology, and that IS

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

therefore can learn a lot from knowledge about basic epistemological positions. Second, that in order to overcome its difficulties, IS must change its orientation towards the family of historically oriented epistemologies.

THEORY, METATHEORY, AND PHILOSOPHY

A theory in IS is a theoretical explanation of information systems efficiency, of user behaviour, of the function of different search elements such as descriptors, citations, titles, and so on. We do not have many explicit theories in IS. Actually, it is difficult to name just one good example. Often theories from other fields, for example, psychology, sociology or management are applied, but they are not theories of IS. Ranganathan's facet-analytic approach contains a theory of subjects, but some authors, including [2], do not count classification research as a part of IS. Also, what is called 'information theory' is by many people – this author included – not regarded as a theory in IS, but a theory in computer science. Specific approaches such as algorithmic retrieval or citation-based retrieval should not be termed theories, but they rest on a basis of assumptions, which can be termed 'metatheoretical'. An example of such a metatheoretical assumption could be that the more times a given term appears in a text, the greater is the likelihood that the paper is about the concept that is expressed by that term. Metatheoretic assumptions are thus broader and less specific than theories. They are more or less conscious or unconscious assumptions behind theoretical, empirical, and practical work. Metatheoretical assumptions are connected to philosophical views, and are often parts of interdisciplinary trends, which again can be connected to the *Zeitgeist*. According to [2] the most important metatheoretical approaches in IS up to now have been the 'physical paradigm' and the 'cognitive approach'.

It is a well-known fact that IS lacks good theories. Most work is of a pragmatic nature, which resists scientific analysis and generalisation. However, a lot of papers are published and much practical work is done without explicating any theoretical or metatheoretical assumptions. This makes it very difficult to do theoretical, historical, and philosophical work in IS. However, behind all kinds of activities are certain assumptions about the world in general, about human beings, about language, cognition, research, and so on. Such assumptions *can* be analysed. Philosophy is the field of knowledge, where most general knowledge of this kind is stored and organised. Philosophy learns from the single sciences, it generalises this knowledge, and it communicates this knowledge back to different sciences in different amounts. IS can learn from philosophy, but philosophy cannot dictate principles to other sciences. There must be co-operation, and IS must understand its own philosophical problems.

The most basic disciplines in philosophy are ontology/metaphysics and epistemology. Ontology and metaphysics are about what exists, about basic kinds, categories, properties, and so on. Epistemology is about knowledge: what knowledge is, how we get knowledge, the basic methods of coming to know, etc.

For reasons of simplicity I classify different epistemological approaches in three main groups: empiricism, rationalism, and historicism. Of course, this is extremely simplified. Concepts currently used in the literature are about postmodernism,

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

social constructivism, neopragmatism and so on. This is a big field that many people study for a lifetime; it has its own journals, databases, and so on. It is also a field making progress, and a field that must be interpreted very carefully.

Empiricism is a philosophy that favours perception and experiences. Rationalism is a philosophy that places less relative emphasis on sensory experience and more on reasoning and *a priori* theorising. These two movements arise, in part, from different ways of drawing epistemological and methodological lessons from the ongoing progress of the scientific revolution inaugurated by Copernicus and consummated by Newton. Together, rationalism and empiricism constitute the two main tendencies of European philosophy in the period after scholasticism and prior to Kant. Empiricism is connected to British thinking, Rationalism to Continental thought, but this classification is a crude one [3, p. ix].

Historicism is a philosophy that emphasises that perception and thinking are always influenced by our language, culture, by our pre-understanding and 'horizon', including our scientific theories. Historicism has a strong connection to the humanities where hermeneutics has been dominating for centuries. As a theory about science, historicism has especially evolved as scientific realism, which is an evolutionary epistemology developed within American pragmaticism (by Charles Sanders Peirce) and within historical materialism (by Friedrich Engels) in the nineteenth century.

The twentieth century has been dominated by empiricist philosophy, especially up to 1950. With the computer revolution came a new rationalist trend, which dominated in the seventies and eighties. Thomas Kuhn [4] is, however, an example of an increasing historical influence in modern philosophy. In the 1990s, historicism seems to have become a dominant epistemology.

Different approaches in IS (such as the physical and the cognitive paradigms) can be understood as parts of more general, interdisciplinary theoretical trends. Each trend has its own strengths and weaknesses, which can be illuminated by a more conscious philosophical analysis. IS has been very much dominated by viewpoints related to empiricism and rationalism, but in recent years more interpretative, historical, and neopragmatic views are beginning to influence the field.

Philosophy can not only be used to analyse the metatheoretical issues in IS but also substantial theories in IS. Indeed, my feeling is that such philosophical problems play a dominant role, and I have sometimes been tempted to say that IS is a kind of applied epistemology. Implicit philosophical assumptions not only lie behind the work of information specialists, but also behind the behaviour of information producers, users, intermediaries, and the traditions in the documentary systems. This kind of theory is really deep.

USERS AND COGNITION

There is a close relation between epistemological theories and psychological theories. Psychological theories about users, their behaviour and their cognition are playing an important role in IS. However psychology itself is currently debating its own theories and their epistemological assumptions. This debate is relevant for the understanding of users and cognition in IS. The basic epistemological theories also represent basic view on users and their cognition.

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

Empiricism saw users as born without any knowledge (*tabula rasa*), and all the knowledge an individual obtained came from the senses. Users form simple concepts from simple sense impressions. By the laws of associations more complex concepts could be formed in the individual. Experiences must always be fragmentary and private. In the twentieth century this view has been carried on in behaviourism, which dominated American psychology from 1913 to about 1965. This view has influenced IS in many ways, most recently and most explicit in theories about neural networks and 'connectionism'.

Rationalism, on the other hand, saw sense experiences as a limited way to attain knowledge. In order to see something, a person must already have a certain psychological make-up, which permits him or her to interpret the sense data. A person must have some concepts and these concepts cannot come from the senses but must be inborn (or they must develop from some preform, which is inborn). In modern terms: the brain must run some programs or follow some rules, which determine the fate of all input and the actions of the individual. This view of psychology dominated 'the cognitive revolution' starting about 1956 with the psychologist Jerome Bruner and the linguist Noam Chomsky, dominating psychology from about 1965, and culminating about 1985. It was closely connected to research in artificial intelligence and to the interdisciplinary field known as the cognitive sciences. Today there is a re-evaluation and discussion about the status of this interdisciplinary trend [5] and many people find its epistemological assumptions very problematic.

Historicism agrees with rationalism in the view that our experiences are determined by our psychological make-up. However, it does not see this make-up as something inborn or common for all human beings, but rather as determined by cultural factors. Where cognitivism compares the human mind with a computer, and tries to explain logical thinking, the working of the memory, and decision making as governed by rules which can be uncovered and used in systems with artificial intelligence, historicism understands psychological mechanisms as culture-determined. 'Logical thinking' in 'developed countries' is opposed to 'wild thinking' in 'primitive cultures'. One explanation is that the development of written language changed the cognitive functions. In cultures with written languages it is possible to compare the formal structure of sentences, whereby formal rules of logical thinking can be formulated and taught. Even members of a literate culture who have not had courses in formal logic will be affected by this new way of thinking [6]. Such a way of explaining logical thinking is very different from cognitivist assumptions. In this way the psychologist Lev Vygotsky (1896–1934) saw higher cognitive functions such as memory as determined by culture. Primitive societies think more in pictures, where developed societies have a more verbal functioning of memory. The memory of small children works by biological principles, but with the learning of a language, memory begins to work on a new higher level determined by socio-cultural factors. This cognitive theory was developed around 1930, but today (in the 1990s) it seems to represent a main stream in international psychology.

In the philosophy of science, historicism has been influential in the work of Thomas Kuhn [4]. His theory about scientific paradigms reflects how the processing of information by scientific knowledge-producers (and users) is determined by

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

more or less conscious assumptions. Kuhn's theory bridges the individual and the collective level in cognitive processes. Kuhn's view did not, however, influence the basic thinking about users and cognition in IS for a long time. The influence of historicism in this discipline was in particular introduced by the criticism which Winograd and Flores [7] posed to traditional rationalism (in the form of artificial intelligence and cognitivism). This implies a new view of users as social and cultural beings, and of a more sociological-epistemological view on information seeking. There are several different schools working under the broad headline of historicism, for example, hermeneutics, pragmatism, social constructivism, semiotics, and activity theory/the cultural-historical school. However, it is beyond the scope of this article to present each of them. My own view is presented in greater detail in [8, 9].

From the historical viewpoint information-seeking behaviour can be understood as governed by epistemological theories or assumptions. This is also the case even in the information seeking behaviour of people who have never been taught epistemology. Peoples' epistemological assumptions can be more or less conscious, shadow-like or contradictory, but they can be interpreted from a horizon established by philosophy and the history of science. There should be no dualism between theories of the information seeking behaviour of the (information) scientist and theories of the information seeking behaviour of the ordinary user. Both parts can have a deep or a superficial understanding of source criticism, of the relevance of pre-understanding, of cognitive authority, of interdisciplinarity, and so on. This perspective provides user studies with a necessary theoretical point of departure and also gives an explanation of the lack of progress and cumulativity in the thousands of user studies performed since their introduction in 1948 at The Royal Society Scientific Information Conference in London.

WHAT IS 'A SUBJECT'? AND HOW SHOULD WE PERFORM SUBJECT ANALYSIS?

I consider subject analysis of documents as one of the most fundamental activities of library and information professionals (including in computer applications). Therefore, a theory about subjects and subject analysis must be very important. This is also the field where I have done most of my work: both analysing other theories and developing my own theory [8, 10]. I shall very briefly introduce the different views and refer interested readers to my book.

The subject of a book (or any other document, or message) is closely related to what kind of answers people can find from reading the book. Different people put different questions in different situations and one single document can in principle answer an infinite number of questions. I define the subject of a document (or of any message or sign for that matter) as the epistemological potentialities of that document [10]. Any document thus has an infinite number of subjects. Subject analysis in IS is the process by which the subjects of documents are determined. Because a document does have an infinite number of subjects, the process of subject analysis is a process of giving priority to those subjects which best serve the needs of the users of the information system in question. This view of subject matter differs basically from mainstream approaches in IS.

One approach (often called 'the physical paradigm' [2]) considers information retrieval as an objective, neutral process, where the solution is a 'technological

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

fix' that can be measured by 'recall' and 'precision'. Algorithmic approaches in IS are based on such thinking and on the presumption that the subject of a document is a function of the words in the document (sometimes even that the subject can be described by extracting words from the document). In other words, the 'subject' is implicitly regarded as a 'semantic condensation' of the document. In my analysis, this view is related to the empiricist view.

Another approach (often called 'the cognitive view' [2]) relates the subject of a document to a user's knowledge (or rather to his or her anomalous state of knowledge). Information is here seen as an object, which can fill a gap in an individual person's knowledge. By using cognitive psychology's study of human information processing, it is imagined that it is somehow possible to build information systems, which can relate the content of documents to individuals' needs. In this way, there is a connection back to a rationalist influence.

My own approach (the 'sociological-epistemological paradigm' or the 'domain analytic approach' [8, 9]) recognises that a given document may serve different purposes for different user groups (or for individuals in different situations). Subject analysis should neither be seen as universalistic or as individualistic, but should reflect the target groups of the information systems. The same document should receive different subject descriptions if represented in say a medical library compared to a military library. The more general subject analysis (e.g. the one made by Library of Congress) should analyse documents from the major theoretical or ideological perspectives in the same way as is done in, for example, the history of literature, the history of science, and the history of ideas. (In principle such an analysis contains the same problems concerning subjectivity of judgement as exist in such interpretative disciplines.) The difference between an analysis made for a specific purpose and an analysis made for a general purpose can be compared to the different perspectives in applied science and in basic science. The specific purpose requires a more pragmatic perspective, while the general purpose requires a more realistic perspective. My view is connected to historicism.

This analysis of the concept of subject demonstrates how different conceptions and assumptions in IS are influenced by metatheoretical and epistemological views. In the next section I shall go a step further.

THE METHODS OF CLASSIFICATION

I can illustrate my claim about the importance of epistemological theories by addressing the problem of how to classify documents, which is both a classical problem and still a core problem in information science. It is very hard to find presentations and discussions of the methodology of classification in the literature of IS. A few approaches, such as facet analysis or factor analysis, do represent a specified method, but there is no analysis of the implicit assumptions behind such different methods and there exists no theoretical framework to compare the relative strengths and weaknesses in different approaches to classification.

It is my opinion that different methods of classification basically reflect different epistemological theories, as shown in Figure 1. The relative strengths and weaknesses behind these approaches cannot be found in the literature of IS, but

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

	Research Objects (Scientific classification)	**Documents** (Bibliographic classification)
Empiricism	Classification provided by statistical analysis (such as factor analysis) based on 'resemblance'. *Examples*: Classification of mental illness in psychiatry or kinds of intelligence in psychology based on statistical analysis of test scores.	Documents clustered on the basis of some kind of similarity, e.g. common terms or bibliographical coupling. *Examples*: 'Atlas of science' and 'Research fronts in *SCI*', algorithms for information retrieval.
Rationalism	Classification based on logical divisions, e.g. classification of people in age groups. *Examples*: Frame-based systems in AI; Chomsky's analysis of deep structure in language; cognitive models of the mind in psychology.	Facet analysis built on logical divisions and/or on 'eternal and unchangeable categories'. *Examples*: Ranganathan, Bliss II and Langridge; semantic networks.
Historicism	Classification based on natural development. *Examples*: The theory of evolution: Biological taxonomies.	Systems based on the development of knowledge-producing communities (the division of scientific labour). *Example*: The feature of DDC that it distributes subjects by discipline.
Pragmatism	Classification based on analysis of goals and consequences (critical classification).	Systems built on critical analysis of the development and state of knowledge. *Examples*: Francis Bacon, the French Encyclopedists, Henry Bliss, the Marxists etc.
Scepticism (including post-modernism)	Ad hoc classifications Unstructured Internet resources as a model	

Figure 1. *Fundamental methods of classification*

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

must be found in the literature of philosophy. This is in my opinion a very strong argument for the relevance of epistemological theories.

In practice, research libraries and classification systems often employ subject specialists to develop and update their systems or they import important parts of their system from recognised handbooks or similar authoritative sources. But this is only to move the problem one step back: how do you know when a given source reflects 'cognitive authority'? How do you distinguish between good and bad proposals? In order to evaluate this you must develop a theory about the methodology of classifying.

Epistemology has no final answer, there is no consensus about *the* scientific method. Insight in epistemology can, however, provide you with knowledge about the merits and weaknesses of the different solutions, and progress in the scientific method as well as in classification must be based on the historical evidence gained in epistemology and science studies.

Figure 1 shows the relationship between basic epistemological theories and basic methods of classifying. Classification is done in all sciences. Like any other science IS has different approaches to classification based on different epistemological views. IS is mainly concerned with principles for classifying documents produced in other disciplines, which implies classification on a second order level (or meta level). Classification in IS is not restricted to documents, but can be applied to all forms of 'information' represented in information systems. Different sciences may influence each other. Frame based systems and semantic networks are examples of classifications developed in AI and also applied in IS. 'Facet analysis' is a method of classification developed independently in IS and in psychology. A science which can export its methods to other sciences is regarded as a stronger science.

In my opinion there exist a limited number of basic methods of knowledge organisation corresponding to basic epistemological views. For example a psychiatrist can classify mental illness using empirical methods, rationalistic methods, historical methods or pragmatic methods (or, of course, combinations of these). In the same way, a psychologist can classify forms of intelligence or mental capacities by using statistical analysis of test scores (empirical method), by using computer models (rationalistic method), by studying the social construction of the intelligence concept (historical method), or by choosing a concept which fertilises his general perspectives and aims.

On another level, information scientists can use the same kinds of methods to organise documents, knowledge, or information. They can use empirical methods, such as bibliometric linking, and produce maps such as the 'Atlas of science'. Or they can use rationalistic methods, such as developing facets or principles for logical division, they can use historical methods such as revealing the cultural bias in different systems, or they can select classifications which support the aim of their activities.

In this paper it is not the job to outline or to argue for my own epistemological view, which is done in other works, e.g. [8]. The whole idea is to demonstrate that these different metatheoretical views still play an important role and that a qualified investigation of them seems mandatory for IS.

Two basic problems in IS are: (a) how independent should bibliographical classifications be in relation to scientific classifications; (b) what epistemological

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

method, or combination of methods, should be used. Traditional ideals have been empirical or rationalistic, providing 'neutral' or 'objective' classifications. Modern epistemology, however, emphasises the theory-laden character of observations, which also imply the theory-laden character of classifications: they are not neutral discoveries but constructions, which favour some kind of activities at the expense of other activities. This implies a movement from more positivistic approaches toward more interpretative and neopragmatic approaches.

COMPOSITION, SEMANTICS AND RETRIEVAL

All information retrieval is based on the matching of search terms to some 'subject access points' which may either be a part of a document itself, or which may represent some kind of value added information provided by information specialists or others. Normally IR applies some algorithms without any theoretical basis in the composition of documents or the specific value of different access points. In this way, traditional IR is very reductionistic. An alternative non-reductionistic approach to IR proposed by the present author [11] is a theory about the value of a given term considering its specific field or place, e.g. the relative value of title-words compared to words from abstracts or descriptors (or the relative value of references compared to term-searching). Research has demonstrated that such relative values vary over time and over knowledge fields, so that no general mathematical function can ever be expected to work equally well in all databases. This insight makes mainstream IR research look very problematic.

On-line searching applies certain heuristic rules, such as: searching the title field provides more precise, but less complete retrieval compared to abstracts and full text [8, pp. 23–25; 11]. However, such heuristic rules depend on the concrete conventions used in creating the field concerned. In the social sciences, titles often use metaphorical language, which can make title searching misleading. Similar problems exist with other fields. In biochemistry, for example, methods are usually cited but not reagents. This indicates that reference searching by means of citation indexes should be a useful heuristic strategy when searching for methods, but not when searching for reagents. On-line heuristics are not independent of content, which indicates that it is necessary to develop a research programme in database semantics, which can provide the necessary heuristic guidelines for retrieving information. Such a research programme must uncover the conventions of sub-languages and genres in different kinds of documents and in different domains.

Semantics is an interdisciplinary field, which studies the meaning of words and symbols. Also in semantics we have many different approaches which, at the deepest level, reflect basic epistemological positions. The young Ludwig Wittgenstein formulated a 'picture theory' of meaning related to the views of logical positivism, whereas the older Wittgenstein formulated a theory of 'language games' much more related to pragmatic philosophy. Much research in IR seems to be based on presumptions more related to the picture theory than to the theory of language games. In a way, many information scientists may regret that Wittgenstein did not develop in the opposite direction because the picture theory seems much more fit for formalised IR than the theory of language games.

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

However, IS should build on the most probable semantic theory, not on some unrealistic or obsolete presumptions, which only fit their dreams. (This is explored in much more detail in [11].)

The theory of language games – as well as related theories developed by pragmatic philosophers like Peirce and Dewey – indicates that the meaning of words depends on their use. We use language as a tool to fulfil certain goals and the meaning of our words reflects both past history and future goals. Concepts are not universal phenomena, linked to the brain, but are shaped in specific social activities and internalised during learning. Such a view of semantics has important implications for retrieval theory. Such a theory must be much more connected to specific domains and their sub-languages.

THE MEANING OF 'INFORMATION'

According to Buckland [12, p.6] the term 'information' is used in different ways in IS, including 'information-as-knowledge', 'information-as-thing' (data, document, recorded knowledge) and 'information-as-process' (becoming informed). According to Buckland information is always situational. What is informative in one situation need not be informative in another situation. I agree with this view, which I develop further. Different documents (or different texts, signs or things) have different meanings in different domains of knowledge, and should therefore be interpreted differently by different information systems [8, pp. 110–112].

Thus, 'information' is used both in the meaning of 'document' and in the meaning of the knowledge transferred by documents. What we today call 'Information Science' was once termed 'Documentation'. One of the most pronounced changes towards 'Information Science' was the decision of the American Documentation Institute (ADI, founded in 1937) to make an official change of its name to the American Society for Information Science (ASIS) in 1968. But this institutional change of name did not solve the theoretical issues about the meaning of 'information'.

Ellis [2, pp. 187–188] describes an anomaly in IS: that computer systems are using Shannon's theory, whereas information retrieval systems (IR) are not based on a measurement of information, but of physical entities (relevant and not relevant documents). 'Brookes [13] noted the anomaly could be resolved if information retrieval theory were named document retrieval theory which would then be part of library science. However, he commented that those working in the field of information retrieval were making the explicit claim to be working with information not documentation'.

I do not agree that those working in the field of information retrieval are making the explicit claim to be working with information as opposed to documentation (or documents). Such a claim seems absurd, and the most influential and modern information retrieval project mentioned in Ellis' book is TREC (Text REtrieval Conferences). In 1988 P. Willett published a book entitled *Document retrieval systems*. These are but two examples demonstrating that information retrieval researchers do not object to working with documents.

However, many researchers in IR share the hope or the ideal that it should somehow be possible to retrieve not only documents, but to retrieve the knowledge

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

or the facts contained in the documents and even to measure the amount of information retrieved. B.C. Brookes certainly belonged to this group of researchers. Such a view is connected to an extreme form of empiricism/positivism and reductionism. My own view is that the ways people are informed are mediated by institutions, by documents, by language and by other cultural products, and that the factual content of messages and signs cannot be isolated from these cultural mediators [8, pp.17–19]. People are not only seeking 'raw facts', but also substantiated knowledge claims. Therefore users are interested in background information, and this is transferred via documents/texts, informal communication and other means.

The above mentioned reductionistic view has had the unfortunate consequence that IS has invested much too little time in studying documents, their typology, composition, and their role in informing users. This is again an important argument for a move away from empiricism and rationalism towards more historic epistemologies.

THE TYPOLOGY OF DOCUMENTS

Today, there exists no general theory of documents in information science. I believe that such a theory must be connected to a theory of communication [14]. Documents are historically developed tools, which are formed in ways which facilitate their purposes: to communicate knowledge, or to store knowledge, which can be seen as communication over time. There are many kinds of communication and documents. Most documents, e.g. newspapers and magazines, are more related to mass media research than to IS. A very broad theory, which has been useful in media research as a point of departure for documents in general is Jürgen Habermas' 'Theory of the public sphere' [15].

Information science has traditionally been more concerned with scientific literature and its retrieval in systems such as MEDLINE or *SciSearch* than with the study of mass media. This is not to say that these fields cannot learn from each other or should not co-operate, but only that their focus has been and still is somewhat different.

One important element in a specific theory of documents in information science is the analysis of functional differentiation between, for example, primary documents, secondary documents, and tertiary documents. Primary documents comprise, e.g. scholarly monographs and scientific articles. Secondary documents comprise, e.g. electronic databases like MEDLINE, *SciSearch*, and OPACs. Tertiary documents include encyclopaedias and review articles. Further functional categories are source documents (e.g. in history) and what has been termed 'repackaged information' such as textbooks and popular science. The last mentioned type represents a fifth functional category designed for the exportation of scientific knowledge to other user groups. Price's [16] famous article about the literature of N-rays made an important contribution to a theory about the division of labour between the primary and the tertiary literature – and thus to the whole understanding of these functional systems.

Another element in a theory of documents is their location in specific domains. Every domain of knowledge has its own more or less specific types (e.g. in music: sheets of music; in geography: maps and atlases; in law: codes/bodies of law; in

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

astronomy: almanacs; in genealogy: pedigrees and genealogical trees and in psychology: tests). Often documents in one domain are inspired from another domain. This is often connected with theoretical influences. When, for example, the social sciences try to follow the methods of the natural sciences, their primary, secondary and tertiary documents also tend to follow the norms from the natural sciences.

This fact leads us to the third element in the understanding: their dependency on epistemological assumptions. Not only do the content of documents, but also their sub-languages, their composition and whole system of documentation reflect epistemological presumptions. The kind of research undertaken in composition studies by C. Bazerman and others [17, 18], provides us with one valuable method to study documents (e.g. experimental articles) from a modern epistemological point of view.

A research strategy in information science studying documents from such a non-reductionistic point of view as outlined above can provide much concrete, detailed and cumulative knowledge of relevance to retrieval systems and services.

INFORMATION SELECTION AND RESEARCH EVALUATION

In addition to classifying, indexing and retrieving documents, libraries and information systems also select documents and maintain collections. This is of course done from the perspective of the concrete library/information system. Libraries for children, public libraries, research libraries, national libraries and databases such as MEDLINE or *SciSearch* select documents using different criteria. Selection development must have both general and specific criteria. What general principles can IS develop concerning document/information selection in information systems?

A collection should be able to provide ‘satisfactory answers’ to the questions raised by actual and potential users. The quality of collection development is related to the ability to meet the requirements of the users and supply them with satisfactory answers.

In the empiricistic and positivistic approaches, the quality of collection development is often based on investigations of the users’ demands and experiences with collections. Data about circulation of different parts of the collection can be used to control its further development. This approach presupposes that what count as ‘satisfactory answers’ is something which user studies can uncover, which implies that it is the users’ subjective experiences of satisfaction, which play the dominant role.

However, what count as ‘satisfactory answers’ can be a result of different trends in knowledge production and consumption. Different approaches exist and can have a more or less dominant role at different times. Knowledge development as a whole is also partly determined by the influences of different trends. The single individual (or a group of individuals, e.g. high school students) can be more or less up-to-date with general norms and tendencies in knowledge production. This indicates that the subjective needs of actual user groups can be a problematic measure of quality, and a more objective indicator of quality should replace such a subjective measure. The problem of identifying objective indicators of quality is related to determining what counts as ‘cognitive authority’, which is really difficult. The

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

question for the single information service is, however, whether to base its collection policy on empirical studies of users, or on the judgement of qualified staff who have the possibility to read reviews, to study the epistemological trends, and to use other kinds of evaluations. Such kinds of evaluations are in the literature often called 'subjective', but the point made here is that so-called objective, positivistic methods can only measure subjective criteria of satisfaction, whereas the so-called subjective methods of interpretation and hermeneutics can uncover more objective criteria of quality in knowledge production.

My suggestion is that the general theory of collection development should be tied to epistemological and sociological studies of knowledge production and use [19]. Information specialists should evaluate the evaluations and be specialists in such issues as the scientific referee process, the review literature, the function of prices and evaluations, the theory of 'paradigms', and how co-operative and competitive relations in different fields might affect quality. Such studies are also done in 'science studies', but in IS the focus should be the application of this knowledge in the management of information systems.

INFORMATION SPECIALISTS AND INFORMATION SYSTEMS

I shall conclude this paper by presenting some lines of research, which, I feel, tend to reorient IS away from empiricism and rationalism, and towards more historicist epistemologies.

Information science is concerned with research, which might help improve the design of information systems and services. The goal of information systems is to improve users' possibility of finding satisfactory answers to the questions they ask of such systems. Information systems collect, analyse, organise, describe and retrieve information/documents in order to inform actual and potential users.

The term 'information system' is a child of the computer age. When using the term, one often thinks of computer-based retrieval systems such as DIALOG. As an afterthought, manual libraries are also considered information systems, as are journals, encyclopaedias, and the whole formal and informal system of scientific communication ('the social system of science'). When considering, for example, a library as an information system, it is usually considered from a more or less specific 'systems theoretical' point of view, e.g. using terms such as 'input', 'output', and 'feedback'.

However, the communication systems of science (and other social systems) are much older than the computer age, and have, through centuries, developed important characteristics such as source criticism, principles of rhetoric, standards for publishing, and so on. All this represents production, dissemination, and use of information, which is the declared object of research in information science. The understanding of this social system is a pre-condition for establishing computer-based systems to make the system more efficient. If this kind of knowledge is lacking systems design might be a mistake. Such a mistake was made when information scientists in the 1960s decided to replace the journal system in psychology with a computer based system for delivering articles according to the individual users' research interests [20].

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

Traditionally, scholars (humanists) have played an important role in research libraries and scientific communication. With the development of computer technology, people from technological fields, and also people from the social, behavioural, and cognitive sciences, began to play an important role and to analyse the problems of communication from new perspectives. In this process, however, new metatheoretical and epistemological theories flourished at the expense of more hermeneutic and historicist oriented epistemologies from the humanities. There have also been tendencies to an overemphasis on information technology (IT) rather than information resources as the object of IS and to underestimate the virtues of traditional communication systems.

One important perspective is therefore represented by those investigations, which analyse the historical developments of information systems as adaptations to specific communicative needs. Such studies can be published under many labels, one of which is 'social constructionism'.

Another important perspective is represented by those studies, which focus on content, and, for example, compare information systems and information seeking behaviour in different domains of knowledge (e.g. science versus the humanities).

A third important perspective is represented by attempts which focus on the functionality of information systems: analyse explicit and implicit functions and values, look at competing information channels from the users' point of view, and investigate the consequences of commercial and non-commercial conditions and values. Related to this third perspective are investigations, which try to diagnose and repair malfunctioning in information systems. This should be a central ability for information specialists, not only to use different information systems, but to provide qualified argumentation about how to improve their quality and efficiency. Such diagnosis of systems can be relatively simple when, for example, the coverage of databases is evaluated, and it can be very complicated as when it is discussed whether the norms of the information systems in the natural sciences are also appropriate in the social sciences and the humanities.

CONCLUSION

This paper has presented a very broad picture of different areas of information science. It is based on a summary of my own theoretical research in IS published elsewhere. I have demonstrated how different views of knowledge in a very profound way affect all important problems in IS. Epistemological theories have a fundamental impact on theories about users, their cognition and information seeking behaviour, on subject analysis, and on classification. They have also fundamental impact on information retrieval, on the understanding of 'information', on the view of documents and their role in communication, on information selection, on theories about the functions of information systems and on the role of information professionals. In all these questions different epistemological positions can be shown to influence research in a very profound way. I have also tried to demonstrate that the most satisfying solutions for IS in my opinion can be obtained by moving away from such reductionistic and fundamentalistic theories as empiricism and rationalism. Instead IS should approach the big family of historic oriented epistemologies.

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

Philosophical knowledge has been very neglected in IS and the epistemological and metatheoretical views have seldom been formulated or analysed. Instead of conscious analysis such views have mostly been unconscious attitudes by information scientists. It is important for IS to raise its theoretical and philosophical level, the better to understand the limitations and possibilities of different approaches. This article is first and foremost an argument for taking theoretical and philosophical studies in IS seriously.

REFERENCES

1. Vickery, B. Metatheory and information science. *Journal of Documentation*, 53(5), 1997, 457–476.
2. Ellis, D. *Progress and problems in information retrieval.* London: Library Association Publishing, 1996.
3. Garrett, D. and Barbanell, E., eds. *Encyclopedia of empiricism.* London: Fitzroy Dearborn Publishers, 1997.
4. Kuhn, T.S. *The structure of scientific revolutions.* 2nd ed. Chicago: University of Chicago Press, 1970.
5. Johnson, D.M. and Erneling, C.E., eds. *The future of the cognitive revolution.* Oxford: Oxford University Press, 1997.
6. Goody, J. *The interface between the written and the oral.* Cambridge: Cambridge University Press, 1987.
7. Winograd, T. and Flores, F. *Understanding computers and cognition: a new foundation for design.* New York: Addison-Wesley, 1986.
8. Hjørland, B. *Information seeking and subject representation: an activity-theoretical approach to information science.* Westport, Conn., London: Greenwood Press, 1997.
9. Hjørland, B. and Albrechtsen, H. Toward a new horizon in information science: domain analysis. *Journal of the American Society for Information Science*, 46(6), 1995, 400–425.
10. Hjørland, B. The concept of 'subject' information in science. *Journal of Documentation*, 48(2), 1992, 172–200.
11. Hjørland, B. Information retrieval, text composition and semantics. *Knowledge Organization*, 25(1/2), 1998, 16–31.
12. Buckland, M. *Information and information systems.* New York: Greenwood Press, 1991.
13. Brookes, B.C. Information technology and the science of information. In: Oddy, R.N., Robertson, S.E., van Rijsbergen, C.J. and Williams, P.W., eds. *Information retrieval research*. London: Butterworths, 1981, 1–8.
14. Brier, S., Dyrby, M., Graulund, J., Koggersbøl Hansen, I., Hjørland, B., Vind, T. and Ørom, A. *Faglitteraturens dokumenttyper i kommunikations – og videnskabsteoretisk belysning. Kategorier, medier, former, genrer, niveauer og kvaliteter.* Bind 1–2. Copenhagen: The Royal School of Library and Information Science, 1997.
15. Habermas, J. *The structural transformation of the public sphere: an inquiry into a category of bourgeois society.* Cambridge, Mass.: MIT Press, 1989. (Translated from: *Strukturwandel der Öffentlichkeit*, 1962)

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib

Journal of Documentation, Vol. 54, No. 5, December 1998

16. Price, D.J. de Solla. Networks of scientific papers. *Science*, 149, 1965, 510–515.
17. Bazerman, C. *Shaping written knowledge: the genre and activity of the experimental article in science.* Madison: University of Wisconsin Press, 1988.
18. Bazerman, C. and Paradis, J., eds. *Textual dynamics of the professions: historical and contemporary studies of writing in professional communities.* Madison: University of Wisconsin Press, 1991.
19. Hjørland, B. et al. *Faglitteratur: kvalitet, vurdering og selektion. Grundbog i materialevalg.* Göteborg: VALFRID, København: Danmarks Biblioteksskole, 1996.
20. Boffey, P.M. Psychology: apprehension over a new communication system. *Science*, 167, 1970, 1228–1230.

(*Revised version received 28 June 1998*)

Aslib, The Association for Information Management
Staple Hall, Stone House Court, London EC3A 7PB
Tel: +44 (0) 171 903 0000, Fax: +44 (0) 171 903 0011
Email: pubs@aslib.co.uk, WWW: http://www.aslib.co.uk/aslib