Peter Jaenecke

To what End Knowledge Organization?[1]

Dr.Peter Jaenecke (b.1943), studied physics and philosophy in Giessen, works at present at the research center of Alcatel SEL, Stuttgart on problems of information retrieval and knowledge management. Interested in cognitive science and philosophy of science. Chair of German ISKO Chapter.

Jaenecke, P.: **To what end knowledge organization?**
Knowl.Org. 21(1994)No.1, p.3-11, 5 refs.
This paper is put up for discussion to help Knowledge Organization find its proper place. To begin with, a few basic concepts from the field of communication theory are defined. The new concepts permit the conclusion that what we are facing at present is not a flood of information but one of messages, and that this latter flood must not be equated indiscriminately with a knowledge flood. Messages must be judged on their contents. To this end, 'knowledge' is subdivided into 'core-', 'peripheral', and 'pseudo-knowledge', and it is argued that the great majority of scientific publications contains peripheral and pseudoknowledge. With the aid of two interlocking and self-amplifying recursive mechanisms it is shown that and how pseudoknowledge is more and more gaining ground, particularly outside the mathematical and scientific disciplines, hence precisely in those fields which deal with social problems in the broadest sense of the word. Thus a deficit of knowledge on which to base action is produced, leading to a general disorientation - a modern form of ignorance, keenly manifesting itself at present in a crisis of leadership. The research tasks resulting from the above for Knowledge Organization are roughly outlined. (Author)

1. Introduction

This paper was originally entitled: „What is Knowledge Organization"? I deemed it urgently necessary to clarify this question, for, even though our Society for Knowledge Organization has been in existence for several years already, we still disagree fairly widely on what 'Knowledge Organization' is. Not that I was after a precise definition, for I doubt a useful one can be found; other disciplines, too, manage quite well without one, so why should precisely Knowledge Organization form an exception? There are formulations in dictionary style such as „Physics is the science of the forms of movement, of the material forces producing them and of their properties", which, while permitting a rough classification of the discipline concerned, can in no way serve as a definition from which consequences can be derived. What I was interested in knowing was what fields belong to knowledge organization and how they are interrelated, hoping to learn in this way just what links the various, rather heterogeneous fields together. However, I was never able to pass beyond a few block diagrams. For I found out quite soon that I lacked a regulatory idea: to be able to name the fields belonging to knowledge organization one must know the objective which knowledge organization pursues. So I shelved for the time being the question „What is Knowledge Organization"? in order to find first of all an answer to the question „To what end Knowledge Organization?"

Pertinent attempts undertaken so far furnish no answer to this question; they concentrate too onesidedly on the improvement of methods. But methods are mere tools: it cannot be the purpose of a discipline to continuously perfect its tools, just as if it were the purpose of physics to build more and more precise measuring instruments. We need a statement oriented to contents, permitting us to reply the questions related to meaning, such as: Why are better classification methods necessary? What problems are they supposed to solve?

Approximmately at the same time when I started to reflect on the purpose of Knowledge Organization I chanced upon critical investigations on the decay of the political parties in which particularly the lack of expert knowledge on the part of political leaders was made respsonsible for the decline of the political party system. Mediocrity rules the day; a political career is hindered rather than furthered by comptence in a given field, for what is demanded is adaptation to the spirit of the age, coupled with an effective self-presentation in the media. By the same token, politicians regard it as a matter of secondary importance to deal with tasks of more than immediate, short-range interest, i.e. precisely with those matters which enable them to give proof of their leadership qualities. This explains their helplessness vis-à-vis many vital problems of our time.

Once my attention had been alerted, I subsequently took a somewhat closer look at the blunders and scandals reported on in the media. I very soon had to realize that the phenomenon of 'incompetence' is widespread not only among the leaders on the political stage, but also among prominent personalities in science and art as well as in the management of business companies, labor unions, churches and sports associations. Recent publications strengthen the impression of a striking leadership crisis extending beyond the borders between institutions or countries.

Such a widespread phenomenon of incompetence cannot be explained by failures on the part of individual persons; rather, there must be a quite general reason for it. Incompetent is the person who cannot fulfill his or her tasks because he or she knows too little about the measures

Knowl.Org. 21(1994)No.1
P.Jaenecke: To what End Knowledge Organization?

3

that need to be taken. Evidently the key to a proper understanding of the crisis lies in the lack of adequate knowledge for action. But what is the cause of such a knowledge deficit? The answer to this question is at the same time the central thesis of the present paper:

The knowledge deficit comes about because the store of knowledge has fallen into disorder so that it is becoming more and more difficult to inform oneself properly on any given field.

Such a store of knowledge which has fallen into disorder is an insupportable and, in our high-tech world, extremely dangerous situation calling for immediate counter measures. This brings me back to my question: To what end knowledge organization? An obvious answer is: to order knowledge, to make it accessible. Now this answer I would like to clarify somewhat by answering questions like the following: What does it mean to say that the store of knowledge has fallen into disorder? What meachnisms cause this disorder and how do they operate? Why have the proposed solutions produced so little effect so far? What tasks does Knowledge Organization have in the search for a solution?

2. Basic Communication-Theoretical Concepts

More than most other disciplines, communication theory suffers from the ambiguity of its basic concepts. It is particularly the concepts 'message' and 'information' as well as the concepts derived from them which keep making for confusion. To prevent any and all misunderstandings, we define here:

A message is a sequence of characters arranged sequentially in space on a medium. A measure for the information value of a message is the extent of the state transitions it produces in a receiver (1).

We distinguish here between the message and the medium on which it is stored, although both belong inseparably together: there are no messages pure and simple, but only objects containing a message, e.g. books, punched tapes, floppy disks, etc.; often they are grouped together under the superordinate concept 'document'[2]. As a sequence of characters, a message constitutes an entity which exists independently of a receiver; information, on the other hand, is always related to a receiver. One and the same message may be differently informative to different receivers, but also to one and the same receiver at different times, for the information which a message provides to its receiver depends of his knowledge and his previous history. If, for example, someone receives a certain message for the second time, this message, although still constituting the same character sequence, may now differ in information from the first time: its informative value may have grown, as the message is now better understandable to him, or it may be near-zero (as in the case of a joke one has heard before). Even evident

nonsense may provide information, for it is not the truth of the message which is decisive, but its effect on the changes of state it produces in the recipient.

On its way from the head of the sender to that of the receiver a messsage undergoes repeated transformations. If such transmission is to fulfill its purpose there must be something which remains invariant under all transformations; this we call the 'contents of a message':

The contents of a message is the sum of all possible information that may be extracted from it.

Unlike information, the contents of a message is something which is independent of the receiver. A confused message is interpreted differently by all recipients, and the set of all possible information comprises particularly many elements here. An exact message has the property that all receivers who have understood it are approximately in the same situation with respect to what has been transmitted; hence they have all received the same information, i.e. the set of all possible information consisting only of this one element. In this special case, provided the message is new to the receiver, contents and information may be identical. Between 'message' and 'food' there is a certain analogy, with the contents of a message corresponding to the nutritive value of a given foodstuff, and its information to what an organism absorbs from that foodstuff.

To use 'information' in combination with 'flood' is automatically ruled out by our definitions: it is not with information, but with messages that we are flooded, and the correct version of the much-quoted communication crisis reads: in the sciences, the quantity of messages is growing exponentially. Whether, however, this high rate of increase of the messages produced goes hand in hand with an equal increase of knowledge, is a different matter. Here it is first of all necessary to clarify the rather confused concept of knowledge.

We place knowledge on the same level as the contents of a message, but not every such contents constitutes knowledge as well:

A message contains knowledge (constitutes knowledge) if its contents consists in universally valid statements on the world.

'World' comprises nature in the widest sense, but also the things created by man, and the universal validity applies in the ideal case to four different levels. The statements must be 1) general statements on a great number of things, not merely on individual things. They must also be 2) ageless, hence not lose their validity after a certain time. They must 3) permit rational decisions of general interest and not merely of interest to a special circle of persons, and they must 4) be exact so that, in principle, every human being may obtain from them the same information.

4

3. Knowledge Misery

Not every message claimed to constitute knowledge meets the above conditions. To be able to better classify messages, we divide them with respect to their contents, into messages providing core, peripheral or pseudo-knowledge.

3.1 Core, Peripheral, and Pseudoknowledge

In mathematics and in the natural sciences, especially physics, almost the entire knowledge is stored in theories regarding as firmly assured. Although some future additions can never be wholly excluded, that which is known already today will be valid in the future as well. Even in 100 years the periodic table of elements or the theory of electrodynamics will not look essentially different from today, and in particular they will not greatly increase in size. Knowledge of this kind meets the above four conditions; we call it *core knowledge*. It is mankind's true treasure of knowledge.

In addition there is, in all sciences, a field of research in which the results are not, or not yet, wholly secure. Here, too, knowledge is produced, but everything is still in flux. Existing knowledge may be expanded, errors may be corrected, or sometimes a new error may even be added. Such knowledge, attractive mainly because of its newness, we call peripheral knowledge. Typical examples are: critiques, discussions pro and con, as well as research reports and academic theses. Peripheral knowledge violates the above conditions 2) - 4): while its truth value can in general be determined, it is of limited generality and can soon become obsolete or be absorbed into core knowledge. But despite its short-livedness, it may render important service at its time, although this service is of an auxiliary nature and will be forgotten when it has completed its task. Peripheral knowledge is only destined for a specific circle of persons, namely the scientists working in the field concerned, and so its importance generally is confined to that field only. Peripheral knowledge may be likened to scaffoldings erected around a building, the core knowledge, to repair or further complete it.

In the past few years there has been an ominous increase in works characterized by 'modern' themes and by statements made with self-assurance and self-proclaimed scientific competence. But what these works proclaim is *pseudoknowledge*. It likewise violates the above conditions 2) - 4), but is neither true nor wholly false, as they link together ideas which, while having a true core, do not belong together. In contrast with peripheral knowledge, which finds firm support in core knowledge, pseudoknowledge is built up wholly out of itself. Pseudoknowledge resembles scaffolding built around other scaffolding or sometimes even wholly in the void. As it is not always simple to recognize the incompatibility of ideas, it is not easy either to distinguish with certainty between peripheral and pseudoknowledge, particularly in disciplines with only little distinctive core knowledge.

Pseudoknowledge is encountered most frequently when a discipline tries to adapt the results of another one.

Pseudoknowledge betrays itself by a few unmistakable characteristics: vague ideas or uncertainties in a field outside one's specialty lead to an *obscure mode of expression*, frequently going hand in hand with an *impermissible usage* of technical terms from that field. In addition, *objective and methodical shortcomings* may be noted, which always indicate that the subject concerned was not properly understood. Further characteristics are: the simplification of problems and, closely related to it, *impermissible generalizations*. Ambiguities can in general be resolved from the context, but in texts containing pseudoknowledge this is no longer possible, as here the context itself is ambiguous or unintelligible.

Pseudoknowledge must not be confused with the snobby terminologies to be found in many scientific publications. Such linguistic facades are usually harmless, recognized and exposed as they are as stupidities. In contrast with pseudoknowledge, linguistic facades can be rephrased into meaningful statements, although here, too, the boundaries are fluid.

An attentive reader will have no difficulty catching a good bag of pseudo knowledge in the contemporary literature. To illustrate this we content ourselves with the following example in which at least three characteristics of pseudoknowledge can be noted:

Circumstances may cause the confirmation degree of a hypothesis the more to decrease „the more empirical proof is furnished for it. From this situation, only evolutionary epistemology can provide a way out by regarding the demanded rationality of the apriori probability as rooted in a phylogenetic aposteriori. The basic principle of this rationality has already been fixed on the genetically conditioned level of the inborn expectation probability metric".

This text makes no sense, although it may be based on a correct insight. What does a phylogenetic apriori have to do with a hypothesis? Is there any sense at all in introducing an empirically determinable confirmation degree for a hypothesis, and what assertive value does it have if it can keep changing? A metric, in mathematics, is a distance measure for two points in a space; the properties of this measure are fixed by axioms. Whatever 'inborn' may mean - an inborn metric is therefore an impossibility. Since reference is made to probabilities, the other meanings of 'metric' from the fields of poetry and music evidently are not meant here, either.

> *We are drowning in a flood of messages and are thirsting for knowledge*   Stiegelbauer

3.2 Flood of Messages and Mountain of Knowledge

To obtain a first impression of whether the high rate of increase of the volume of messages goes hand in hand with a similar increase of knowledge one would have to determine the respective shares of core, peripheral and

Knowl.Org. 21(1994)No.1
P.Jaenecke: To what End Knowledge Organization?

5

pseudoknowledge in the overall message production. This would require very detailed and careful studies surpassing, of course, the capabilities of any single person, apart from the fact that suitable evaluation criteria are still lacking so far. The following is therefore just a personal impression based on my reading experiences, but an impression shared by several other authors.

When leafing through an old scientific dictionary or handbook one will be surprised at what was already known so long ago; hence, knowledge cannot have increased as much as the mass of present-day publications seems to suggest. Nevertheless, every generation contributes something to the store of knowedge, thus ceaselessly increasing it. To characterize the phenomenon of mass in the field of knowledge we will not speak of a flood of knowledge but of a *mountain of knowledge* incorporating all existing knowledge. Even if it has not increased in the same measure as the number of publications, its height has meanwhile likewise become a problem.

But here, too, differentiation is needed. If there are several types of knowledge, the mountain of knowledge cannot be a monolithic block. According to our afore described subdivision it consists, figuratively speaking, of a hard inner core surrounded by a diffuse, moldable peripheral layer - the core and peripheral knowledge, both obscured by a dense mist of pseudoknowledge. To judge by the number of publications, the scientific successes of the present should be greater than ever before. But on closer inspection such data prove to be deceptive (2): Core knowledge is presented in relatively few handbooks only, while by far the major portion of publications pertains to peripheral and, increasingly, to pseudoknowledge. A large part consists of copies, reprinted or repeated ad infinitum in different versions by different authors. This gives rise to an immense, often redundant flood of messages[3].

The flood of messages and the mountain of knowledge are two independent phenomena; they have different causes and require different treatment each. Pseudoknowledge is screened out, while peripheral knowledge ages or becomes part of core knowledge, thus making it possible to keep the quantity of knowledge within limits. The publications in which it is presented, on the other hand, are preserved for all times. There is no remedy against the flood of messages, but it is absolutely in our hands to keep knowledge overseeable. The question: To what end knowledge organization can therefore be answered as follows:

The task of Knowledge Organization should consist in contributing to processing the store of knowledge in such a way that it becomes once more overseeable for man with his physical limitations.

This objective is of a general nature; now it must still be proven that the objective is worthwhile. How important is the overseeability of knowledge? This can be clarified best by examining the consequences resulting from non-overseeable knowledge. They come to light through the recursive mechanisms of knowledge acquisition.

4. Recursive Mechanisms in Knowledge Acquisition

Knowledge acquisition takes place recursively by way of various interlaced and self-amplifying cycles. Each cycle consists of several intermediate stations which are passed through successively. The results obtained in any completed cycle serve as the input data for the next cycle, hence the self-amplifying effect. When left to themselves, such cycles lead either to a continuous, usually imperceptible improvement or to a deterioration. Whether the self-amplifying effect makes matters better or worse depends in large measure of the quality of the input data. The message flood and the large amount of pseudoknowledge make one fear that matters will rather tend to get worse. In the following we will consider two cycles which confirm this assumption.

4.1 The Message Flood Intensifies the Message Flood

In the first cycle the message flood promotes specialization. More specializiation makes for an increasing urge to overcome the isolation connected with it, which results in more and more dilettantism and incompetence. The incompetence produces pseudoknowledge, and pseudoknowledge, finally, again increases the message flood.

This phenomenon is only too well known: there are too many publications. In his lifetime, man cannot read very much more than 2000 books. A simple numerical example will illustrate how small this quantity is in comparison with the quantity of books produced: an avid reader will on the average manage to read one book per week, making for 50 books a year or 1000 books in 20 years. In Germany alone some 100,000 new books appear every year; to read them one would need 2000 years. During this time, with the production rate remaining constant, some 200 million further new books would appear in Germany alone. This comparison, which includes neither foreign books, nor articles in periodicals, shows how ridiculous a misproportion there exists between reading matter volume and reading capacity. Since physically man can absorb only soandso many messages, while the messages keep increasing incessantly, he must, in order to maintain an overview, continuously narrow his field of work. Knowledge fields are abandoned, not for having become uninteresting or unimportant, but because there is no time to give attention to them. When the number of publications continuously increases while the subject matter discussed in it pertains to a more and more specialized field, one's mental horizon must necessarily become narrower and narrower. Nevertheless, publication goes on and on, for even in the case of a trivial subject matter man tends to go to the limit of his abilities. Each new publication adds to the store; further specialization is the result, and at the end of this cycle we find a specialist who

6

Knowl.Org. 21(1994)No.1
P.Jaenecke: To what End Knowledge Organization?

knows everything about nothing: *the message flood boosts specialization.*

The specialist cannot be blamed for his attitude, for it is an act of self-preservation, arising as it does from his thoroughly understandable wish to maintain his competence. But in this he will succeed only if, yielding more and more territory, he withdraws to a further and further reduced field, i.e. for his competence in a smaller and smaller field he must pay by becoming incompetent in a larger and larger one. Incompetence means: In disciplines outside his own, and may be even in matters very close to life, a specialist, for lack of relevant knowledge, can no longer contribute to the discussion or take rational decisions: he loses his faculty of discernment outside his special field: *progressive specialization makes for more incompetence.*

Most people make their peace with their specialist status; a few, however, have recognized their situation as unsatisfactory and try to break loose from their specialisthood by turning to a further, an additional discipline. Now, however, precisely because they are specialists in one discipline, they are laymen in the other ones and therefore of necessity have to familiarize themselves with their new field. Actually this should not be possible at all, for we have explained the trend toward more and more specialization with the steadily growing amount of subject matter, which forces one to limit his or her special field. Very soon they will come up against their physical limits. For lack of time, but also because of their incompetence, it will hardly be possible for them to penetrate to the core knowledge of a foreign discipline. Their domain remains that of peripheral knowledge obtained from secondary literature and mixed with pseudoknowledge. Superficiality is the price they must pay for being able to cope with the additional mass of material. Superficial readers, however, become superficial authors who produce pseudoknowledge rather than knowledge. *Incompetence causes pseudoknowledge.*

Pseudoknowledge lowers the quality level, causing the number of those to increase who feel called upon by their newly acquired knowledge to have their say, and since therefore most of them publish, *pseudoknowledge increases the flow of messages.* This applies particularly to such fields as epistemology and psychology, which, touching as they do on problems accessible to people's own fields of experiences, convey a feeling of competence - but the message flood is likewise boosted by so-called scientific journalism directed at subjects of interest to the media.

4.2 Pseudoknowledge Produces Pseudoknowledge

Pseudoknowledge, being mainly concerned with subjects from the world of life, will, in a further cycle crossing the path of the aforementioned one, lead to new pseudoknowledge. In this (likewise self-amplifying) cycle, pseudoknowledge increases the incompetence, the incompetence leads to disorientation, the disorientation produces malcommunication, which, in turn, promotes the formation of pseudoknowledge.

Pseudoknowledge produces the illusion of familiarity with the subject, thus preventing one from coming really to grips with a subject and penetrating to the core of knowledge. Pseudoknowledge produces a craving, satisfied by numerous publications, for uncomplicated material confirming the prejudices created by the pseudoknowledge already digested and making one receptive to further pseudoknowledge. Loosely interconnected highlights will be the only knowledge 'pads' accumulating in the brain. Simultaneously, the relationship to reality is getting more and more lost, and a pseudoworld is built up which is communicated by the limited experiences gathered. This pseudoworld, steadily fortifying itself and thereby aggravating the incompetence, greatly narrows the selection of messages received: messages which could help to break up the pseudoworld are no longer perceived: *pseudoknowledge increases the incompetence.*

> *We are continuously being informed, and are disoriented nevertheless.* K.Steinbuch

He who lives in a pseudoworld will sooner or later get into conflict with the real world and thus find himself confronted with insoluble problems: with one's personal store of knowledge becoming infused with a larger and larger proportion of pseudoknowledge characterized, among other things, by being neither wholly false, nor wholly true, people become more and more incompetent, disoriented and uncertain in their judgment; in their distress, they reach out to any surrogate ideal offering itself to them; they become more susceptible to political and social ideologies, to occult doctrines of salvation and - in the sciences - to questionable paradigms: *incompetence leads to disorientation.*

Disorientation means also that the ability has been lost to properly assess and classify one's own work. A great many papers read at scientific meetings of every type are therefore merely monologs; no one is seriously interested in other people's arguments, as is proven by the innumerable fruitless discussions which frequently serve only as pretexts for presenting one's own point of view, with misunderstandings being the unavoidable results. A Babel of tonges of truly Biblical dimensions threatens to come about: disorientation produces malcommunication. In his novel „Klim Samgin", GORKIJ has described this process with great mastery.

True communication brings divergent points of view closer and closer together as the discussion continues, with an understanding being reached in the end. Malcommunication, on the other hand, has the opposite effect; the longer it lasts, the more the people concerned will be talking at cross-purposes: *malcommunication makes people garrulous and promotes the production of pseudoknowledge.*

Knowl.Org. 21(1994)No.1
P.Jaenecke: To what End Knowledge Organization?

7

4.3 Disorder in the Store of Knowledge

Our lack of knowledge of ourselves, so the medical scientist CARREL recognized as far back as 60 years ago, does not come from any difficulties in procuring the necessary basic knowledge or in obtaining more than sparse and imprecise data. On the contrary: it is precisely the tremendous abundance and disorder of the store of knowledge assembled by mankind on itself in the course of time (3) which is to be blamed here, with the very abundance of the material causing us the greatest trouble in arriving at its correct and useful application[4]. CARREL was thinking here particularly of the practical application of medical knowledge, but the two mechanisms have made it clear that the sciences themselves, too, are less and less able to make use of their own results: core knowledge remains nearly unchanged, while the highly technical peripheral knowledge of the specialists and the pseudoknowledge of the amateurs keep proliferating more and more.

The store of knowledge has become disordered by the mass of peripheral knowledge widely scattered throughout innumerable publications and the increasing quantity of pseudoknowledge.

The peripheral knowledge of the specialists is not yet utilizable, or at best only technically so, while the pseudoknowledge of the amateurs reveals itself as irrelevant: in a time of a growing need for action, the knowledge produced is less and less suitable for solving vital problems, and despite overflowing stacks at the libraries, ignorance steadily increases. This seems to be a contradiction, but what matters is not what is in the books, but what the knowledge people have available in their heads. Here, it seems, the only choice we have today is that between specialisthood and dilettantism - both frequently united in one person, with the one being as unsatisfactory as the other: a specialist loses his competence outside his special field, while a dilettante has ever had any competence to begin with. Both, for lack of suitable knowledge, frequently are no longer able to have a say in vital matters: they have become incompetent. Mental paralysis, helplessness, lack of far-sightedness, wrong decisions and mismanagement are the unavoidable consequence, as can be read daily in the newspapers. Evidently it is becoming less and less possible to cope with the problems of the present, let alone those of the future.

The two recursive mechanisms of overall knowledge activities leave nothing good to hope for in the future. If they remain active unchecked, the destruction of the store of knowledge and, with it, the slow but unavoidable decay of our culture is inevitable. In view of these prospects, knowledge organization appears to be an urgent task of great social importance. What should its field of activity look like?

This concluding question will occupy us in the next section.

5. Knowledge Organization

Since several decades, justified complaints are regularly being voiced on the message flood and its harmful effects on society. Strange enough: although every scientist suffers massively from the message flood, all proposals on how to end or at least curb it have so far remained wholly without effect; everyone knows this evil and suffers from it, but hardly anyone does anything against it or is willing to exercise somewhat more restraint in publishing. Another peculiarity typical of our time is that, disregarding the true causes and speaking from a too narrow point of view shaped by personal negative experiences, people rashly propose solutions directed against individual symptoms such as increasing incompetence, loss of values, malcommunication in the sciences, etc. Moreover they do not examine whether the proposal is realizable or whether, if implemented, it would really improve matters, with the effort frequently even being shunned to prove whether the proposal has in fact anything to do with the shortcomings criticized.

Such proposal, although made with the best of intentions, are already clearly marked by incompetence: as the totally unusable concept apparatus proves, the illness proceeding from the message flood has meanwhile affected also the persons willing to fight it. The fact is usually overlooked that we are dealing here with a complex cybernetic system whose partial systems continuously adjust themselves to one another, so that individual measures have hardly any chances of success. Quite generally, too, it should not be left to laymen to look out for suitable countermeasures, as if such measures were already known and need only to be applied. This is by no means the case; rather, discovering such measures should be a research task for Knowledge Organization.

Its main objective has so far been the ordering and supply of knowledge. Thus understood, knowledge organization in fact always comes too late, for it lets the chaos-producing forces have their way and contents itself with clearing-up operations. But knowledge organization must, in its work, be concerned with the future as well; it must exert its influence on many wrong developments identified so that disorder may be prevented. That means: it must both cope with the mass of existing knowledge and contribute something to its reduction. Pursuing these two aims at the same time makes very good sense if only for the reason that in many cases they require the same methods. From this there result various partial tasks based upon each other, such as: literature supply, classification of scientific works, and their systematic representation.

5.1 Literature Supply

The methods and objectives in literature retrieval at present hardly take into account the limitations of human capacities. Interest is centered, rather, on the problems of system technology as produced by the flood of publications. The supreme principle adhered to is: Any material

produced must also be kept available. As the mass of data increases, the required technological efforts therefore of necessity increase as well[5]. The decisive question is, however: available for whom? With the improved performance he strives for, the system developer would like to offer better retrieval to the user. But according to what yardstick is 'better' decided upon? When the user physically is no longer able to process the retrieval results because of their volume, any further increase in performance will hinder him in his work rather than help him.

Conditions force us to reconsider things. Technical perfection will only then be useful to the user if it goes hand in hand with relevant data material. In information retrieval a document is considered relevant with respect to a given query if it deals with the subject described in the query (4). This point of view, which is typically that of the system developer, only considers whether or not the document matches the query, but is unconcerned with the contents of the documents. When, however, a query causes thousands of 'relevant' documents to be turned up, the concept of 'relevance' has lost all meaning for the user. Now, although the flood of publications might be curbed somewhat by administrative measures, it cannot be prevented as such, and therefore one should limit at least the flood of documents retrieved. To the user, relevance is inseparably connected with overseeability; it can be achieved by limiting the material on the one hand and seeing to it on the other hand that the search responses will be overseeable.

Simple arithmetic tells us that the hit rate in retrieval will be the smaller, the fewer titles are known to the system. One of the objectives to be pursued should therefore consist in excluding from inclusion into the store as many superfluous documents as possible, for there is no need at all to blindly accept any and all publications. The majority of today's publications deserves to be thrown away rather than to be subjected to serious keyword analysis[6]! Purposeful selection would provide a less drastic means for consigning superfluous documents to oblivion. This presupposes, however, suitable evaluation criteria for the contents of a document, which criteria still need to be drafted. But first and foremost it is the author himself who should have the possibility to withdraw an earlier publication from his hand and replace it by a better one.

While scrupulously exact indexing and the inclusion of irrelevant data material are often the causes of a non-overseeable response to a query, the blame is frequently put on the querier and his allegedly inept query. But the user proceeds in his search from the point of view of contents; he is in search, figuratively speaking, of an earthern vessel and is supplied instead of a box of shards that are not even certified to have come from the same vessel. How, under these circumstances and forced to apply the primitive means of expression of Boolean logic, is he to limit the quantity of shards in a sensible way? Once superfluous documents have been collected, there will alway be queries for which these documents will be

relevant from an information retrieval point of view, but not to the user. Suppose there exist 100 identical documents by different authors. With respect to a suitable query all 100 would be relevant, but for the user only one; he has no way to exclude all the other ones by a more ingenious query. The same will happen to him with useless documents that have not explicitly been marked as such in indexing.

If an archeologist cannot characterize his findings by any intrinsic properties, he will do so by using external characteristics such as: place of finding, type of material, color, shape, size, etc. and will number them for clear-cut identification. In a wholly similar way it is tried to classify literature by assigning keywords. Is this a description according to intrinsic or to external characteristics? The answer is: both the one and the other, depending on whether the publication concerned is to be regarded as a shard or a complete vessel. Formerly, when the publications still more or less resembled entire vessels, the keywords were still to a high degree descriptive of the contents, whereas at the present time, with its peripheral knowledge, it is the shards which dominate, and the keyword, even though it may actually occur in the text concerned, has no ordering function surpassing that of e.g. the color in the case of shards. The loss of quality of the material to be described evidently has had no effects on the practice of content description: keywords are still assigned as they were before, but just as it is more troublesome to describe shards of pottery than complete vessels, so it takes more effort to assign keywords to peripheral knowledge than to core knowledge, while pseudoknowledge, with its empty shells of words that lack consistent contents, actually cannot be sensibly described at all. The poorer and the more confused the content is, the more difficult will the assignment of keywords be; in the end the descriptors will characterize only the message itself, hence in a sense only the wrapping material, but no longer the contents. This development promotes the flood encountered in retrieval.

In information retrieval, when viewed from a user's point of view, two main mistakes are being made that call for greater attention in the future: 1) The document store is not kept free of pseudoknowledge, so that shards and useless stones are granted full, equal status, and 2) a publication, e.g. a book, being a physically self-contained unit, is always simultaneously regarded as a self-contained unit from the point of view of content, i.e. shard and vessel are placed on the same level.

5.2 Assessing Scientific Documents

Acquiring relevance does not only mean to present as thorough as possible an overview of the entire document store, it also means avoiding the superfluous and not letting disorder come into existence in the first place. It is of these two partial tasks that the classification of a scientific document consists. The first one requires its fitting into the set of documents already available and

concerns chiefly the librarians, while the second one, requiring as it does a quality evaluation, mainly concerns authors and publishers: A suitable set of evaluation instruments is still a desideratum[7]. So far there have been only few research projects in which texts themselves are the objects of investigation. We therefore still have too little experience in the desciption of the contents of texts and the recognition of structures of thought, particularly with a view to machine analysis. In the following these two partial tasks will be briefly outlined.

He who has a precise overview will be able to oversee even a large number of things. But 'overview' is a relative concept; there are various points of view and manners of viewing, and therefore it will often be necessary to satisfy highly disparate requirements. A student of the history of science, e.g. will also be interested in the errors of a given epoch and therefore would hate to miss the documents containing pseudoknowledge, the very ones which the specialist in the given field would like to have excluded. Contents description, proceeding descriptively as it does, can solve this conflict only inadequately, for it concentrates on the description of contents, while what matters here is the identification of the relationships existing between a new document and those already available in the system, e.g. whether something already known is disproved, supplemented or merely repeated. We are dealing here with knowledge about the store which does not occur in the documents themselves. It can preferably be presented as a semantic network and enables the user to develop his own manners of viewing and employ these in searching for literature. Thus, if a „vessel manner-of-viewing" is employed, the statement *'shard x is a part of vessel y'* permits a much better selection than the description of the form and color of this shard, and statements like *'x contains the same as y'* permits a much more effective limitation of the retrieval flood than the inclusion of further descriptive searching concepts.

Just why are so many superfluous papers being printed? Evidently this happens because the publishers, who meanwhile likewise have fallen victim to the general loss of competence, succeed more and more rarely in fulfilling their control duties. But why do the authors go to so much trouble to write superfluous papers for which they usually are not even paid? Publihers and authors alike seem to lack a quality yardstick. The quality of scientific papers might already be noticeable improved by obligating the authors to adhere to certain minimum requirements. These, however, will be obvious demands of a purely formal nature, e.g. that the relationship to other publications should be indicated, whereas quality control in the true sense of the word must be based on the state of the art in the given field, which is nowhere better represented than in a theory conceived of as a systematic representation of a subject field.

5.3 Systematic Representation

Scattered over many basically poor documents one may find numerous individual good thoughts; hence it is not only the quantity of knowledge as such which causes us trouble, but also its parcelling up into small bits and their elaborate wrappping. MERTEN has listed e.g. 160 „definitions" of communication (5). After reading all of them one might expect to be perfectly informed, but in fact one will find onself more confused than ever before, for they form a collection of individual unconnected aphorisms. CARREL seems to have been the first one to recognize the necessity of a systematic representation:

„Our task consists in making a rational selection from the mass of dissimilar material"[8] „If our knowledge is to be of any use to us it must be available in a concise, synthetic form"[9]

In a systematic representation, several such aphorisms or fragmentary thoughts are taken together and formed into a new knowledge unit according to a specific principle, just as individual shards are combined to make a vessel. The advantage is obvious: the synthesis enables us to weed out non-authentic material, for when it is known what form of vessel must result it is also possible to decide whether a given fragment belongs to it or not. The number of elements is thus reduced, making for better overseeability. In the same way as in a marking process, many individual parts are combined into a new object of greater complexity. The shards lose their identity; their description can therefore be dispensed with. No knowledge is lost in this process, on the contrary, new knowledge is even produced, for the complete vessel is more than the sum of its shards. If the contents of the aforementioned 160 definitions of 'communication' could be compressed into one single coherent thought, these definitions might all be consigned to oblivion without any harm being done.

A not inconsiderable part of the present misery in science can probably be blamed on a questionable research postulate which demands that new discoveries be made all the time. Evidently the production of new, spectucular shards is more meritorious than the combination, in many years of painstaking minute work, of old shards into a usable vessel. The results of this one-sided evaluation of scientific achievement are plain for all to see: scientific work is becoming less and less effective and its results are becoming less and less relevant to society. Evidently we need to revise not only our conception of the type of the scientist, but also the conditions under which scientific research is to be carried out[10]: the discovery of something new must no longer be the sole objective worthy of being pursued by scientific research; rather, an equally high esteem should be accorded to the synthesis of (individual) scientific results.

6. Resumé

The business of science has become questionable, with the sciences mainly producing knowledge for their own private use. Despite a growing need for it there is less and less knowledge to guide our actions, with a general disorientation being the result. This situation should be a challenge to Knowledge Organization. So far it has been

10

concerned particularly with the supply of literature, or, to formulate it rather drastically: while publications were classified and ordered, little attention has been paid to the disorder in the mass of publications, the so-called information crisis. In view of the pressing problems of our time, for which no one seems to feel really responsible[11], this one-sided orientation should be abandoned. Understood thus, knowledge organization comprises activities carried out in single branches of science in the form of disciplines engaged in the production, representation, processing, and utilization of knowledge, including those rendering assistance to this end; in addition to these it comprises methodical themes transcending the boundaries between disciplines, as well as connections to epistemology and the cognitive sciences. The task of the Society for Knowledge Organization should consist in coordinating the various activities of the individual fields concerned.

When there is talk of a shortage of resources one thinks of raw materials and energy. That spiritual resources are just as needed for our survival is something that has so far been overlooked. Just as in other fields, we are living here, too, at the expense of the future: we produce, but do not enrich; we do not create a new tradition, but, through the mental garbage we leave to future generations, we deny them access to the old one. Something must happen soon; things cannot go on (for long) as they have so far.

Notes:

1 Paper given at the German ISKO Chapter Conference, Weilburg, 26-28 Oct.1993.

2 A recorded signal, e.g. the tape-recording of a speech, likewise constitutes a document. The difference between signal and message will not be discussed here.

3 See (2) p.193

4 See (3) p.10

5 Thus e.g. an information retrieval system, is regarded among others as the more effective, the larger its recall (= number of relevant responses to a query divided by the number of all possible relevant responses) is. A system developer will therefore strive to come as close as possible to the upper limit of 100%.

6 This tendency will still be intensified by the worldwide interlinking of computers and the ascent of multimedia documents: authors who even have trouble with the single medium 'text' will have still more trouble with multimedia means of representation. Moreover: a poor piece of work will remain poor even if embellished with pictures and sound. In principle, everyone will be able in the future to disseminate such products worldwide within seconds. A foretaste of things to come is offered today already by Internet News.

7 Electronic publishing may open up wholly new avenues here: The user might be given a possibility to evaluate (perhaps anonymously) the document requested and read by him. Such generally accessible evaluations would be of assistance to future readers, but also to the author himself, who might furnish corrections after the fact or defend himself against unwarranted criticism, none of which he can do in the case of printed contributions.

8 See (3), p.41

9 See (3), p.10

10 See (3), p.58

11 Knowledge organization is a stepchild of scientific research. While in the European research projects RACE and ESPRIT the development of new communication technologies (that will only increase the message flood) is being promoted on a large scale there seems to be no project going on at the moment which is coming to grips with the consequences of the message flood; the Deutsche Forschungsgemeinschaft (German Research Association) at least, according to its own statements, is not sponsoring any such project. Evidently the social importance of this problem has not been recognized at all so far.

References

(1) Steinbuch, K.: Maßlos informiert. Die Enteignung des Denkens. München/Berlin: Herbig Verlagsbuchhandlung 1978. p.52, 55

(2) Oeser, E.: Das Abenteuer der kollektiven Vernunft. Evolution und Involution der Wissenschaft. Berlin u.Hamburg: Verlag Paul Parey 1988. p.153, 189 ff

(3) Carrel, A.: Der Mensch, das unbekannte Wesen. Stuttgart/Berlin: Deutsche Verlagsanstalt 1936. p.41 (Transl.of Man - the Unknown.)

(4) Salton, G., McGill, M.J.: Introduction to modern information retrieval. Auckland etc.: McGraw-Hill Book Co. 1987. p.163:

(5) Merten, K.: Kommunikation. Eine Begriffs- und Prozeßanalyse. Opladen: Westdeutscher Verlag 1977. p.168-182

Dr.Peter Jaenecke
An der Mauer 32, D-75334 Straubenhardt

Knowl.Org. 21(1994)No.1
P.Jaenecke: To what End Knowledge Organization?

11