Robert Fugmann
Hoechst AG, Frankfurt*

On the Practice of Indexing and its Theoretical Foundations

Fugmann, R.: On **the practice of indexing** and **its theoretical foundations.**
In: Intern. Classificat. 7 (1980) No. 1, p. 13–20

This article introduces a series of eight papers giving a survey of the decisions to be made in the choice of the most expedient indexing method and presenting a short description of the contents of each of the papers, namely: The definability of the requested information. – Individual vs. generic concepts: The predictability of their mode of expression. – The fidelity of concept representation: The indexing language vocabulary. – The indexing language grammar. – The category-controlled interplay between vocabulary and grammar in an indexing language. – The practice of information supply: the personal file. – The employment of a large indexing language vocabulary. – The employment of an indexing language syntax. I.C.

"Nothing is more practical than theory".
Ludwig Boltzmann

A. Introduction to the series of papers

In everyday life and in his profession the human is strongly dependent on how fast and reliably he is supplied with information and is able to retrieve it on demand. In all his decisions he must build as far as possible on knowledge and experience which others have accumulated and made available in print or by verbal communication. This naturally saves him the time-consuming, costly, and arduous process of acquiring the experience himself, and in this way he gains access to completely unexpected information which he never would have thought of seeking. But what is more, by compiling, surveying and analyzing very many, seemingly unrelated, individual reports, he can also discover lawful relationships which had previously gone undetected. The resulting gain in knowledge and cognition always provides powerful support for any kind of practical work. Thus one's knowledge of physical laws eliminates much trial-and-error experimentation which would otherwise be required before a workable design for a new aircraft, bridge etc. could be found. It is inherent in cognition that it originates from experience which has previously been carefully reviewed and ordered in an ingenious, new way. This requires memorizing or retrieving existing relevant experience as completely as possible.

All over the world great effort is being spent on disseminating information and making it retrievable and, hence, reusable. It is an especially great problem if the task is to extract a predefined piece of information from a very large store of many kinds of reports and messages which have accumulated over the years. An embarassingly large variety of methods is used for this sort of "directed supply of information", even in one and the same field and often even for apparently the same purpose. These methods are not only markedly different with respect to their effectiveness and cost, they also exhibit very large differences in their indexing languages and their potential for meeting future demands.

A theory of information supply and indexing could prevent many time-consuming, expensive and discouraging failures, if it were able to predict the future behaviour of an information system. Several attempts to deal with this problem have already been reported in the literature (1–15).

Such a theory will be particularly sound if it is based on a small number of general axioms, the validity of which is obvious and in no need of proof. The conclusions and predictions which are derived from these axioms will necessarily be extremely reliable (cf 16). A first draft of such an approach to the theory of information supply was submitted by the author in 1972 (17). Since that time the explanatory value of this theory has been tested against a large sample of cases from both personal experience and published reports. Discussions with fellow members of the Society for Classification* have also contributed much toward further development of the theory.

The ideas embodied in this theory have to a large extent already been expressed by other authors, especially by Rush and Landry (5), Harmon (18), Bernier (19) and, most prominently, Ranganathan (20–22) and his Indian school (23), though in widely scattered publications and embedded in other contexts. An essential aspect of our theory is that it appears to require only the set of five axioms stated below to describe and explain all currently known phenomena in information supply (cf. (24)).

These 5 axioms are as follows:

1. Axiom of definability
   The compilation of information relevant to a topic can be delegated only to the extent to which this topic can be *defined*.
2. Axiom of order
   Any compilation of information relevant to a topic is an *order-creating* process.
3. Axiom of the sufficient degree of order
   The demands made on the *degree of order* increase as the size of a collection and/or the frequency of the searches increase.
4. Axiom of predictability
   The success of any directed search for relevant information hinges on how readily *predictable* or reconstructible are the modes of expression for concepts and statements in the search file.
5. Axiom of fidelity
   The success of any directed search for relevant informa-

* Wiss. Dokumentation, Postfach 80 03 20, D-6230 Frankfurt 80
Deutsche Fassung beim Autor erhältlich.

* Gesellschaft für Klassifikation cV, (Society for Classification), Woogstr. 36a, D-6000 Frankfurt 50

https://doi.org/10.5771/0943-7444-1980-1-13 https://www.inlibra.com/de/page/agb - Open Access -

tion hinges on the *fidelity* with which concepts and statements are expressed in the search file.

A theory promises to be particularly valuable, if it can provide a basis on which to evaluate indexing methods and indexing languages, since these tools are in fact the conceptual core of any system for directed information supply. The consequences of mistaken decisions which have been made on inadequate theoretical grounds are especially far-reaching. The purely empirical approach does not reveal until much too late that one has entered a blind alley, from which there is no escape except by retracing one's steps and beginning again from the start, whereby at least a large part of the fruits of the previous work is lost.

This article introduces a series of eight papers, giving a survey of the decisions to be made in the choice of the most expedient indexing method and presenting a short description of the contents of each paper. It is intended to assist all those who wish to devise an indexing method or design an indexing language for a particular constellation of circumstances and requirements, or who have to evaluate the suitability of an available system for a specific application. The last three chapters in this series demonstrate how the theory has contributed toward the designing of three very different operational information systems. It is also hoped that these papers will lead to a still more advanced theory of information supply in that our theory is confirmed or modified and improved, or completely superseded by a superior theory of even greater explanatory and deductive power. Much will have been achieved, however, if these papers give added impetus to and are able to intensify the discussion on the interplay between theory and practice in the area of information supply.

The following presentation will proceed according to the outline provided by the flow chart on the opposite page. The chart summarizes the decisions which must be made in designing or selecting an indexing method and language to meet current and future information demands, in so far as the latter are foreseeable, in conformity with available technical and manpower resources. Each group of closely related topics will be dealt with in a separate chapter. Chapter summaries are presented immediately below. All technical terms which could be misunderstood are defined in the appended glossary.

Part B (synopsis): The definability of the requested information

The success of any kind of theoretical and practical work depends very largely on the extent to which information about the project at hand is available to the person undertaking the work. In part, this information requirement is felt as a lack of knowledge; the person knows or assumes that information pertaining to his endeavour has been published and he can make this information the object of his seach.

If this search is delegated to an information scientist or, under certain circumstances, to a suitably programmed search mechanism, then the type of information wanted must be communicated in great clarity. This *"directed search"* for information can obviously locate only those documents which correspond to the a-priori definition of the subject of interest. Consequently, for the information seeker much depends on how clearly and completely he can formulate his request (cf. axiom of definability).

Another large part of the useful information is obtained via a different path. It is encountered unexpectedly, as in the reading of technical literature or newspapers or through hearing papers read at a conference. This kind of information could not have been defined beforehand or made the object of a directed search since, previous to its discovery, the seeker was totally unaware even of the possibility of its existence. Its value and usefulness are recognized only subsequent to its discovery and perhaps only after it has brought new understanding. Hence this type of information is accessible only by way of the *"undirected search"*. It is important to realize that both the directed and the undirected ways of supplying information are required in order to satisfy the entire information demand of the human.

Under certain circumstances the requester would accept as a response to his query a type of information which does not satisfy all his – well considered – search requirements. This holds for documents which convey important additional details. Thus, a document on corrosion phenomena *in general* may be of interest to the requester although his subject of interest is the corrosion of copper pipes for drinking water or the intercrystalline corrosion of aircraft wings, if this document provides a general review of corrosion phenomena (and hence may lead to the specific information being sought) or treats these phenomena from the safety or economic viewpoint. This kind of information demand, which lies conceptually between the defined and the undefined demand, can in principle be satisfied by the technique of "reverse retrieval" (40).

In practice many requests turn out to contain search requirements which are inherently indefinable. Then, no one but the requester can judge whether these search requirements are satisfied by a particular document, and consequently he should not object when documents are included with the information supplied to him which do not satisfy his subjective, undefined search requirements, and which he would therefore consider as irrelevant.

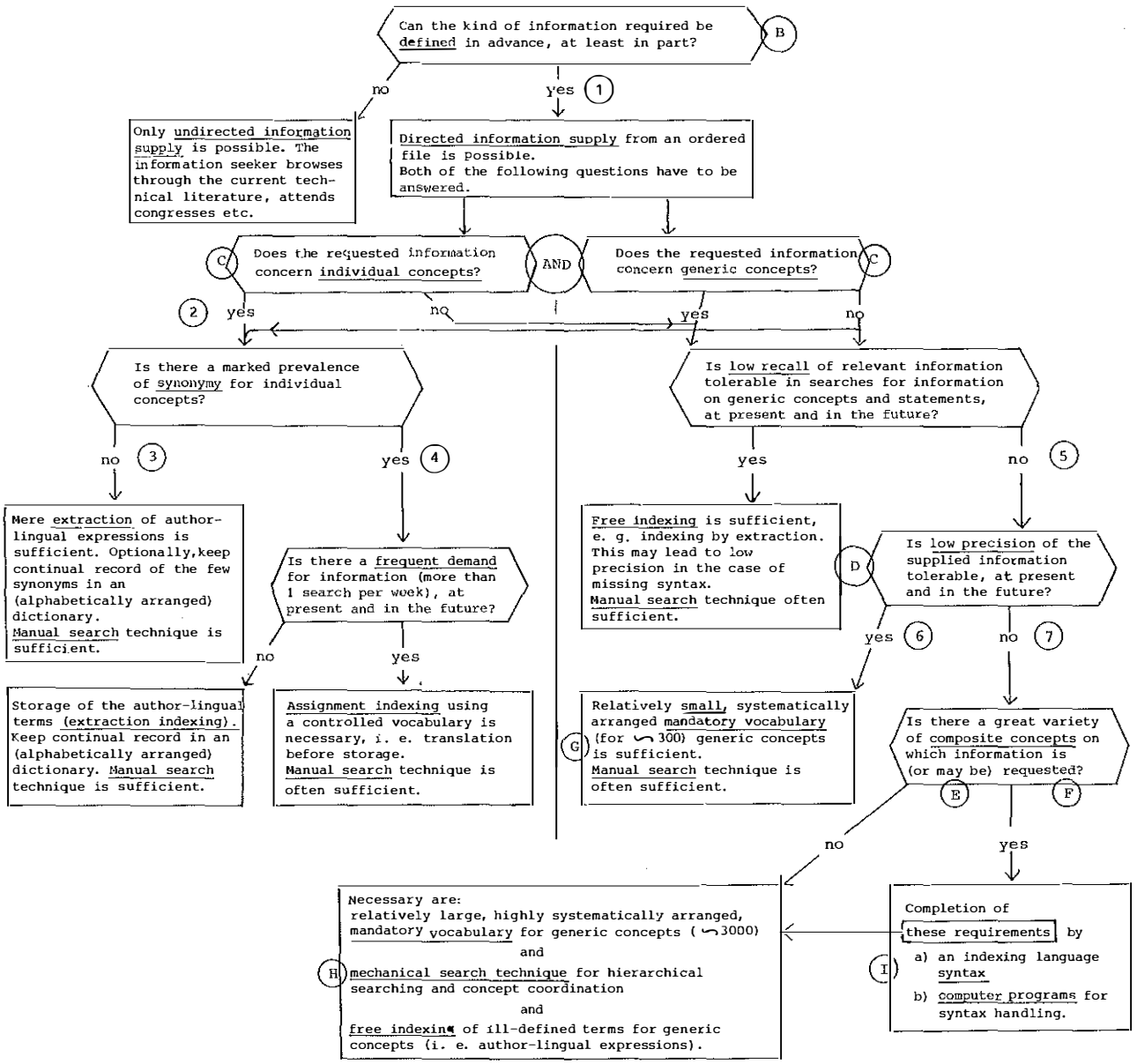
These considerations should prevent requesters from expecting too much from an information system, and they should also prevent information scientists from trying to satisfy an undefined information demand, as such efforts are doomed to failure right from the start. These considerations also reveal the subjectivity of the concept *"information"* and the hopelessness of any attempt to define this concept in "objective" terms. Admittedly, the latter kind of definition is justified in the technology of message transmission, but it would be seriously lacking in usefulness for the field of information science. It will also become apparent that for judging the performance of information systems the concepts of *"relevance"*, *"pertinence"* and *"usefulness"* must be carefully defined and delimited with respect to one another.

Any directed search for information can be viewed as an order-creating process performed on a search file (cf. axiom of order). Several pecularities in the behaviour of information systems can be understood on the basis of the order concept as defined in the natural sciences, especially in thermodynamics.

A phenomenon which is frequently observed in the

https://doi.org/10.5771/0943-7444-1980-1-13 https://www.inlibra.com/de/page/agb - Open Access -

Fig. 1: Flow chart of decisions to be made in the choice of the most expedient indexing method.



Paper B The definability of the requested information.
" C Individual vs. generic concepts: The predictability of their mode of expression.
" D The fidelity of concept representation : The indexing language vocabulary.
" E " " " " " : The indexing language syntax.
" F The category-controlled interplay between vocabulary and grammar in an indexing language.
" G The practice of information supply: The personal file.
" H " " " " " : The employment of a large indexing language vocabulary.
" I " " " " " : The employment of an indexing language syntax.

1 If order in the file is to be achieved on demand and mechanically (retrieval) the documents in the file must have been previously indexed.

2 The author-lingual expressions for individual concepts are readily predictable, often experts are not required for their indexing. Algorithmic (automatic) indexing is promising.

3 For example, in the case of authors (publishing or being cited) institutions, geographical concepts.

4 For example, in the case of chemical compounds, pharmaceuticals, species of organisms.

5 Here, high predictability of the stored modes of expression for generic concepts and statements is necessary, i. e. mandatory indexing, at least for a core of particulary important, categorial concepts. Merely controlled indexing is insufficient.

6 In the case of a file that is (and remains) small ($< 10^4$ documents) and/or when the file is searched only rarely (< once a week) and/or when high topicality and/or extremely low input cost is of paramount importance even if owing to this decision the survival power of the entire system becomes questionable.

7 In the case of a file that is (or soon becomes) large ($> 10^4$ documents and/or when the file is (to be) searched frequently (> once a week).

https://doi.org/10.5771/0943-7444-1980-1-13 https://www.inlibra.com/de/page/agb - Open Access -

search for information is that in the course of time, as the file of documents expands, the search is rendered increasingly more difficult until it becomes almost impracticable. The same can happen to the vocabulary of descriptors if it develops parallel to the document collection. Here the degree of order that prevails in both kinds of collection becomes less and less adequate to meet the demands that are increasingly being placed on it. Oddly enough, this situation develops even though one has merely continued the mode of operation which had proved fully adequate on a small scale.

An extremely high degree of order can be achieved if the file is ordered only on demand and then in such a way that it is tailored to the requester's specific interest profile (to the extent that this profile is definable). If this ordering operation is performed mechanically, i.e. by means of a retrieval operation, then the documents in the file must have been previously *indexed*. Otherwise a high degree of order cannot be achieved in this file or, in other words, retrieval cannot be highly accurate.

Part C (synopsis): Individual vs. generic concepts: The predictability of their mode of expression

To carry out an accurate directed search for information one must know in advance the mode of expression that was used to represent the target concept in the search file, as each of these expressions must be taken into account in the phrasing of the query. Otherwise one runs the risk of loosing relevant information that is contained in the file (cf. axiom of predictability). Viewed in this way "indexing" is the process of

a) identifying the essence of a document to be filed and
b) representing this essence in a mode of expression which is sufficiently predictable or reconstructible*.

That this mode of expression should also be sufficiently undistorted will be discussed in parts D and E (The fidelity of concept representation).

If the requester is content with searching for information pertaining to the *individual concepts* exclusively, then the task of designing and operating an information system is greatly simplified. In this case the natural-lingual mode of expression, as chosen by the author himself, is already sufficiently predictable or can be rendered predictable by simple dictionary routines. This makes it possible to use correspondingly simple indexing methods. Even programmed, automatic indexing methods can serve this purpose fairly effectively.

The author-lingual mode of expressions for *generic concepts*, however, is in most cases not predictable or reconstructible after it has been stored and lost from view. The same holds true for *statement*-sentences with which the author has expressed concept relations of the more global kind (informemes), since the number of conceivable natural-lingual expressions for these concepts and statements is almost unlimited (cf. also (39)). Therefore, if all the expressions, more or less arbitrarily chosen by the authors, are entered into the search file, the task of phrasing a query for a generic concept or a statement becomes extremely difficult or even impracticable, especially if the information supplied is required to be anywhere near accurate.

* Of course, the reconstruction of the modes of expression for concepts in the file must not require any knowledge of the wording of the corresponding document texts, for these are the ones still being sought.

One solution to the problem is to translate the essence of the documents to be filed into a highly predictable indexing-lingual mode of expression. Another method of achieving predictability is to enter the author-lingual expressions into the search file while at the same time keeping in a thesaurus a complete record of the modes of expression for generic concepts and statements already contained in the file.

Both methods require the analysis of each document by an expert. Here the prospects for the use of programmed, automatic indexing techniques are poor if the directed search for information is to be sufficiently accurate. There may be some prospects of success only where extraction indexing is adequate, e.g. in subject fields with great terminological ambiguity, and where exclusively individual concepts are involved.

Automatic indexing always requires specific instructions to be laid down for all possible future document texts. Since these cannot possibly be foreseen and taken into consideration in their entirety, corresponding instructions can also not be provided in a machine program of adequate effectiveness.

Another problem closely connected with generic concepts and statements is that of developing an *indexing language* that can cope with both present and future demands.

In the past the differences between individual and generic concepts, which are crucial from the information science viewpoint, have often been disregarded. The result has been that frequently techniques have been recommended for the retrieval of information on generic concepts and statements merely because of their simplicity and low input cost, although they are in fact suited only for individual concepts. On the other hand, those information systems which have been developed and operated especially to handle generic concepts and statements have frequently been critized because of their, allegedly, excessive expense and complexity of operation.

Part D (synopsis): The fidelity of concept representation: The indexing language vocabulary

The practice of representing the essence of a document in an indexing-lingual mode of expression is always subject to the risk of omitting essential details contained in the original document for the two reasons discussed below. If later a requester is interested in precisely these details they are not available as search requirements. The result is a large proportion of irrelevant documents among those supplied in response to a request precisely for these details (cf. axiom of fidelity). If the file continues to grow and, concomitantly, the frequency of searches also grows (which is highly desirable for the effectiveness of the system), then supplying relevant information from such a file becomes increasingly laborious and expensive. Then searching these files has often had to be restricted more and more until in many cases it has had to be discontinued altogether (cf. axiom of the sufficient degree of order).

The first requirement to counteract such a breakdown is the possibility of representing the details of a search request in an indexing language expression. This, however, is only a necessary but by no means sufficient condition for obtaining a precise information supply. These details must also actually have been expressed in the

https://doi.org/10.5771/0943-7444-1980-1-13 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

indexing language *during the indexing of the documents which contain them.* Otherwise an appropriate search descriptor could not be satisfied by the relevant documents, although it was introduced into the vocabulary precisely for this purpose and is even suggested by the vocabulary for the phrasing of the query. This unfortunate situation always results if for some reason the indexer does not use the full expressiveness of his indexing language. Under these circumstances, the fidelity of an indexing language, as displayed by the high specificity of its descriptors, proves to be only deceptive. At the same time the predictability of the indexing-lingual mode of expression is also impaired. Loss of relevant information and irrelevant responses are the result.

Thus, and as a second requirement, an indexer must not only use the indexing language as a controlled vocabulary, he must also endeavour always to find and use the *most appropriate*, specific descriptors provided by the language for the indexing of a document, since the same choice will also be made for the later phrasing of the query. This *"mandatory indexing"* makes greater demands on the indexer's expertise and requires the structure of the vocabulary to be sufficiently transparent. In particular it should clearly display all the hierarchical and important non-hierarchical relations of the concepts represented in it. Otherwise the indexer could not rapidly and reliably enough locate the descriptors which are most appropriate for indexing a document. Classical logic and the work of Ranganathan and his Indian School provide ample assistance in designing well-structured vocabularies.

Another problem is presented by terms the meaning of which is not entirely clear but about which information is frequently requested, nevertheless. Translating unclear terms into another language always presents difficulties, and translation into an indexing language is no exception to this rule. Furthermore, a pre-established vocabulary cannot provide descriptors for concepts which will emerge or become important only at some future time. For these reasons technical terms of the natural language should also be permissible for describing the essence of a document in addition to the descriptors of an indexing language. They should also be available as search requirements, even though the accuracy of a query employing these terms may frequently be inferior to that of a query employing only vocabulary descriptors. Where individual concepts are concerned, it is advisable in most cases to represent them by natural-language terms.

Whether the pre-established vocabulary should consist of notations from a classification system or of a thesaurus of natural-language terms or some combination of both, will be determined by the prevailing situation and circumstances under which the information system must work.

Part E (synposis): The fidelity of concept representation: The indexing language grammar

If, for the sake of fidelity, a separate descriptor is provided for *each* generic concept or statement, failure will be the likely result. Unreliability will gradually creep into the indexing procedure because a vocabulary of this nature will grow rapidly and continuously; it will nevertheless permanently lag behind practical requirements since "new" generic concepts and statements will occur in almost every new document. The majority of them will be composite concepts which are composed of conceptual constituents already represented in the vocabulary. These precoordinate descriptors will constitute the nodes of a constantly growing and branching network of concept relations in the vocabulary. It will become increasingly difficult for the indexer to follow this semantic network so as always to arrive at the most appropriate descriptor for a generic concept, which he, however, must do to perform reliable, mandatory indexing. This holds true at least for everyday indexing practice, for which only limited resources of time, concentration and memory are available to the indexer for the description of each generic concept.

If in a particular subject field composite concepts dominate, then it is preferable not to represent them exclusively by means of the vocabulary device, but to employ in addition grammatical-syntactical devices; in principle a composite concept is analyzed into the conceptual constituents which are already (or expected to be) contained in the vocabulary.

An indexer may merely enumerate these conceptual constituents, without expressing the pattern of connectivity according to which they are combined with one another to represent the generic concept of a document. But such a low-fidelity representation would mistakenly respond to an entirely different combination of conceptual constituents in a query. Much irrelevant material will inevitably be included in the responses to such a (likewise non-syntactical) query.

If one has an efficient syntax at one's disposal and also the computer program for handling it, then a maximum of fidelity, predictability and, consequently, retrieval accuracy is attainable, since the vocabulary of this indexing language will always remain conceptually transparent and applicable in the mandatory mode. However, considerable effort and great expense are required for the development of such a highly sophisticated indexing language. This expenditure is justified only if there is an assured demand for a highly accurate information supply.

Part F (synopsis): The category-controlled interplay between vocabulary and grammar in an indexing language

In an analyzing indexing language of the kind described above the partitioning of tasks between the vocabulary and the syntax of the language must be carefully controlled. Otherwise, the indexing language mode of expression would not be sufficiently predictable. The semantic concept categories which are encountered in any subject field can serve as a guide for this control. Maximum order in the vocabulary and, consequently, a maximum degree of mandatory indexing are attained if in principle only "unicategorial" concepts are permissible in the vocabulary. These concepts are composed of only one single categorial constituent; any additional conceptual constitutents must be non-categorial. This principle, however, results in much use being made of syntactical devices, and this may entail relatively high input expenses.

If pragmatic exception-rules are sought in order to counteract this effect, then the primary aim should be to control the size and rate of expansion of the semantic network in the vocabulary. Then the vocabulary will

https://doi.org/10.5771/0943-7444-1980-1-13 https://www.inlibra.com/de/page/agb - Open Access -

always remain sufficiently transparent and sufficiently well-suited for mandatory indexing.

Part G (synopsis): The practice of information supply: The personal file

The lowest possible demands on the accuracy of the information supply and on the definition of the technical terms are encountered in the case of a personal file. Normally, here financial and personnel resources are also at a minimum. As an example, an information system for a special kind of document and for special information requirements is described. This system has been in operation for many years, during which it has been developed to maturity. It is demonstrated which conceptual and technical devices have proved to be necessary or dispensable in order to establish and maintain the operancy of the system under the special circumstances prevailing in this case.

Part H (synopsis): The practice of information supply: The employment of a large indexing language vocabulary

In many subject fields a sufficiently high fidelity of concept representation can be achieved and maintained solely through the use of an indexing language vocabulary. It should be sufficiently specific (and correspondingly large) and capable of always being employed in the mandatory mode. High demands are made on the systematic structure of such a vocabulary and on its hospitality towards newly emerging concepts and those which are ill-defined. This calls for the possibility of complementary free indexing. Concept coordination as a device for enhancing retrieval precision is also indispensable in this case. As an example, a vocabulary of the IDC system and its employment in mandatory indexing are described.

Part I (synposis): The practice of information supply: The employment of an indexing language syntax

Particularly high demands are made on the accuracy of the information supply from a file for the world's literature in a large subject field. The IDC system is used as an example to demonstrate how the requirement of the theory are met in this case. It is inherent in an information system for chemistry that the technical concepts are unusually well-defined and that a particularly large demand for information of various kinds is encountered. The exceptionally large number of composite concepts occurring in chemistry necessitates indexing languages with a highly developed syntax. The employment of various syntactical tools is demonstrated, in particular TOSAR and the relation indicators which were specifically developed for the IDC system.

Glossary of terms used in this treatise

Accurracy: The accuracy of the information supply is determined by the extent to which the noise of irrelevant information and the loss of relevant information is avoided. Highly accurate information supply is characterized by high ratios of precision and recall.

Concept: Entirety of true statements that can be made on a subject referred to.
Each of these statements contributes a conceptual feature of the concept. (According to Dahlberg (41)).

Concept, generic: a concept with respect to which at least one more specific concept exists which is meaningful from the perspective of the experts in a subject field.

Concept, individual: a concept with respect to which no more specific concept exists which is meaningful from the perspective of the experts in a subject field.

Controlled vocabulary: Indexing language vocabulary, the terms of which are the only permissible ones for the storage of those concepts which are represented in the vocabulary. The choice of the most appropriate terms of the vocabulary is not obligatory (cf. mandatory indexing).

Hierarchical relation: Kind of concept relation that exists in a system of generically related concepts. Two concepts are in generic relation if they share at least one (essential!) conceptual feature and if one of the concepts has at least one additional conceptual feature.

Indexing: Indexing is representing the essence of a document into an indexing-lingual mode of expression.

Indexing language: The task of an indexing language is to represent concepts and statements with a sufficient degree of predictability and fidelity. Any kind of language that serves this purpose is an indexing language.

Indexing by extraction: A kind of indexing in which the author-lingual modes of expression are used to represent the concepts in the search file.

Indexing, mandatory: Mandatory indexing is a kind of indexing in which it is obligatory for the indexer to use the most appropriate mode of expression that is available in the indexing language (cf. controlled vocabulary).

Information: any message or, more precisely, any "informeme") that proves to be of interest to the addressee. He may have been able to specify *in advance* the kind of message desired. The value of a message may also have been recognized only a posteriori, i. e. only after the addressee has encountered it.
This definition comes closest to those of Shreider (25, 26) and that of Haendler (14). Shannon's notion (27) appears inadequate in the field of information science because this definition disregards the meaning of the message and concentrates on the frequency of occurrence (2, 28–31, 26). There is, however, no lawful relationship between frequency and meaning.
Yovits' notion of "information" constitutes a considerable improvement on that of Shannon because he introduced "uncertainty" and, hence, subjectivity which is inherent in the concept of information (32). Both the removal and the increase (!) of uncertainty provide (positive and negative) information. – Uncertainty does not, however, include the case that messages of an entirely unexpected kind may constitute genuine information. The recipient has never felt uncertainty before he encountered it. An obviously inadequate definition occasionally encountered is one which considers information as any kind of message that merely reduces uncertainty on the part of the recipient.

Information supply, directed: the kind of information supply, in which the selection of documents to be

https://doi.org/10.5771/0943-7444-1980-1-13 https://www.inlibra.com/de/page/agb - Open Access -

scrutinized by the inquirer is delegated, e.g. to an information scientist who acts as an indexer or search assistant. The directed information supply is based on a more or less exact, predetermined definition of the search topic.

Information supply, undirected: the kind of information supply, in which the inquirer himself selects the documents of his interest from a body of documents which has not (or at most only very superficially) been thematically preselected.

Informeme: The information contents of a statement (according to Diemer (33, 34)).

Order: the meaningful proximity of the parts of a whole at a predeterminable place (35). Anyone who requests order to be established in such a "whole" (e.g. in a collection of documents or descriptors) will have to specify in advance what kind of arrangement will be considered by him as "meaningful". The "place" may be a distinct location in a shelf of books or the result of a retrieval operation, e.g. a printed list of references or a stack of cards bearing these references. "Proximity" means that closely related items are close together and non-related ones are kept outside the predeterminable place.

Pertinence: Any message that proves to be of interest ranks as "pertinent" and, therefore, as information. It need not necessarily satisfy the requirements of a search request. Thus even irrelevant information may turn out to be pertinent. Since it is impossible for an inquirer to define his entire field of interest, he cannot expect an information system to supply him with all and only those messages contained in the system which are (or may prove to be) of interest to him (cf. Relevance).

Precoordination or precombination: Precombination is the coining of a descriptor the meaning of which comprises at least two concepts that are also separately (i.e. through descriptors of their own) represented in the vocabulary. An example of a precombining (or "composite", "precoordinate") descriptor is "copper pipes", if "copper" and "pipes" are also contained in the indexing language vocabulary.

Query: Formulation of a search request in the language of the information system.

Relevance: Any message ranks as relevant if it contains all the concepts and concept connections of the inquiry in the desired or higher specificity. The occurrence of the required concepts and connections must be capable of being clearly recognized by an expert in this field, irrespective of the mode of expression chosen by the author, and in particular irrespective of whether they are expressly stated or only implied. Requirements that prove undefinable will have to be omitted from the relevance judgment, because in principle it cannot be expected that somebody other than the inquirer himself can handle them to his satisfaction. A relevant response may or may not rank as an interesting one for the inquirer. It may already have been known to him or he may recognize it as being useless for several reasons (36–38).

Reverse retrieval: variation of retrieval in which, under certain circumstances, information is supplied which is more general than the query or at least of equal specificity. – The literature data are phrased as queries, and the topic of the requester is entered into a termporary file (40).

Statement: Sentence in which at least two concepts, expressed in separate terms, are put into a mutual relation.

Syntax: Device for representing associative (i. e. non-hierarchical) relations between at least two lexical units or descriptors.

References:

(1) Borko, H.: Toward a theory of indexing. In: Inform. Process. & Management 13 (1977) p. 355, (Review article).
(2) Dow, J. T.: A metatheory for the development of a science of information. In: J. Americ. Soc. Inform. Sci. 28 (1977) p. 323.
(3) Slamecka, V.; Gehl, J. (Eds.): Information Science at Georgia Institute of Technology. – Theory of systems. In: Inform. Process. & Management 14 (1978) p. 324.
(4) Shera, J. H.; Cleveland, D. B.: History and foundations of information science. – The search for theoretical foundations. In: Annual Review of Inf. Sci. & Technol. 12 (1977) p. 260.
(5) Landry, B. C.; Rush, J. E.: Toward a theory of indexing II. In: J. Amer. Soc. Inform. Sci. 21 (1970) p. 358.
(6) Jonker, F.: Indexing theory, indexing methods and search devices. New York: The Scarecrow Press 1964.
(7) Jones, K. P.: Towards a theory of indexing. In: J. Doc. 32. (1976) p. 118.
(8) Artandi, S.: Machine indexing, linguistic and semiotic implications. In: Amer. Soc. Inform. Sci. 27 (1976) p. 235.
(9) Maron, M. E.: On indexing, retrieval, and the meaning of about. In: J. Amer. Soc. Inform. Sci. 28 (1977) p. 38.
(10) Heilprin, L.: Mathematical model of indexing. Documentation Assoc., 1957. AD-136-477.
(11) Salton, G.; Yang, L. S.; Yu, C. T.: Contribution to the theory of indexing. In: Rosenfeld, I. L. (Ed.): Information Processing 74. Congress, Stockholm, Aug. 5–10, 1974 (IFIP). North Holland Publ. 1974, and
Salton, G.: A theory of indexing. Regional conference series in applied mathematics. Society for industrial and applied mathematics. Philadelphia, PA (1975).
(12) Stokolova, N. A.: Elements of a semantic theory of information retrieval. In: Inform. Process. & Management 13 (1977) p. 227.
(13) Kochen, M.: Principles in information retrieval. Los Angeles: Melville Publ. Co. 1974.
(14) Haendler, H.: Vorstellungen über die Funktion von Dokumentationssprachen in Informationsverbundsystemen. In: Mitt. Ges. Bibl.wes. u. Dok. Landbau (1975) No. 22, p. 21.
(15) Williams, J. G., Kim, Ch.: On theory development in information science. In: J. Americ. Soc. Inform. Sci. 26 (1975) p. 3.
(16) Svenonius, E.: Good indexing: A question of evidence. In: Libr. Science with a Slant to Doc. 12 (1975) Paper D.
(17) Fugmann, R.: The theoretical foundation of the IDC system. In: ASLIB Proc. 24 (1972) p. 123 (Treatise I).
(18) Harmon, G.: Information need transformation during inquiry. In: Proc. Amer. Soc. Inform. Sci. 7 (1970) p. 41.
(19) Bernier, Ch. L.: Correlative indexes VIII. Subject indexing vs. word indexing. In: J. Chem. Doc. 2 (1962) No. 3, p. 117, and preceeding papers of this series.
(20) Ranganathan, S. R., Gopinath, M. A.: Prolegomena to library classification. London: Asia Publishing House 1967. 3rd ed.
(21) Ranganathan, S. R.: COLON classification edition 7 (1971): A preview. In: Libr. Science with a Slant to Doc. 6 (1969) p. 193.
(22) Ranganathan, S. R.: The COLON classification. New Brunswick, N. J.: Rutgers, the State University, 1965.
(23) Gopinath, M. A.: Classification research in India 1968–1973. Bangalore: DRTC 1974. FID/CR report 14). FID publ. No. 405.
(24) Fugmann, R.: Toward a theory of information supply and indexing. In: Intern. Classificat. 6 (1979) p. 3 (Treatise VI).

https://doi.org/10.5771/0943-7444-1980-1-13 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

(25) Shreider, Iu, A.: Informatsiia i metainformatsiia. In: Naučno-Techn. Inform. Ser. 2. (1974) No. 4, p. 3. This work was discussed in:
(26) Belkin, N. J.: Information concepts for information science. In: J. Doc. 34 (1978) p. 71.
(27) Shannon, C. E., Weaver, W.: The mathematical theory of communication. The University of Illinois Press 1949.
(28) Artandi, S.: Information concepts and their utility. In: J. Americ. Soc. Inform. Sci. 24 (1973) p. 242.
(29) Maniez, J.: Information concepts. In: J. Doc. 34 (1978) p. 244.
(30) Smetacek, V.: Information and communication. In: Inform. Process. & Management 15 (1979) p. 173.
(31) Artandi, S.: Information concepts. In: J. Doc. 35 (1979) p. 92.
(32) Yovits, M. C., Abilock, J. G.: A semiotic framework for information science leading to a quantitative measure of information. In: Proc. Annual Meeting Amer. Soc. Inform. Sci. (1974) p. 163.
(33) Diemer, A.: Informationswissenschaft. In: Nachr. Dok. 22 (1971) No. 3, p. 108, 111.
(34) Haendler, H.: Selection-oriented indication of subject fields and propositions. In: Intern. Classificat. 2 (1975) p. 25.
(35) Leaning on a definition for the concept of order attributed to the philosopher Driesch (he did not include the "predictable place").
(36) Rees, A. M.: Semantic factors, role indicators et alia. In: ASLIB Proc. 15 (1963) p. 350, 358.
(37) Foskett, D. J.: A note on the concept of "relevance". In: Inform. Storage and Retrieval 8 (1972) p. 77.
(38) Kemp, D. A.: Relevance, pertinence and information system development. In: Inform. Storage & Retrieval 10 (1974) p. 37.
(39) Wellisch, The cybernetics of bibliographic control: Toward a theory of document retrieval systems. In: J. Americ. Soc. Inform. Sci. 31 (1980) p. 41, esp. p. 48.
(40) Fugmann, R.: Reverse retrieval: Toward analogy inferences by mechanized classification (Treatise VII on retrieval system theory). In: Intern. Classificat. 6 (1979) p. 85.
(41) Dahlberg, I.: Ontical structures and universal classification. Bangalore: Sarada Ranganathan Endowment for Library Science 1978. 64 p. (p. 17).

saur

Commission of the European Communities
Directorate-General for Scientific and Technical Information and Information Management
on the recommendation of the Agricultural Working Group

The complete thesauri are presented each in four monolingual volumes and one multilingual index volume.
Each monolingual volume contains
– the full thesaurus in the relevant language, sorted alphabetically on all descriptors and non-descriptors and displaying, for each entry, the relational structure in a "word block"
– a systematic section, in the relevant language, presented by microthesauri. Within each microthesaurus, descriptors are arrayed in complete hierarchies (or tree-structure) only, starting from the most generic descriptor and going down to the most specific one. This Systematic section is included to permit a rapid overview of the thesaurus coverage.
The index volume contains the linguistic equivalents of all descriptors, arranged in quadrilingual alphabetical lists sorted successively on each of the four languages. The structure is not displayed. Microfiches displaying descriptors in trilingual, structured arrays are located in a pocket at the back of the index volume.

Agricultural Economics and Rural Sociology – Multilingual Thesaurus

5 Volumes ISBN 3-598-10097-3 (the set) DM 200,– £ 48.– $ 111.–

1. **Deutsch** XX, 80 p. ISBN 3-598-10098-1 DM 80,– £ 19.– $ 44.–
2. **English** XX, 84, p. ISBN 3-598-10099-X DM 80,– £ 19.– $ 44.–
3. **Français** XX, 84 p. ISBN 3-598-10100-7 DM 80,– £ 19.– $ 44.–
4. **Italiano** XX, 78 p. ISBN 3-598-10101-5 DM 80,– £ 19.– $ 44.–
5. **Quadrilingual Index** VIII, 130 p. and 16 microfiches ISBN 3-598-10102 (not available separately)

Food – Multilingual Thesaurus

5 Volumes ISBN 3-598-10103-1 (the set) DM 300,– £ 71.– $ 167.–

1. **Deutsch** XX, 132 p. ISBN 3-598-10104-X DM 120,– £ 29.– $ 67.–
2. **English** XX, 148 p. ISBN 3-598-10105-8 DM 120,– £ 29.– $ 67.–
3. **Français** XX, 144 p. ISBN 3-598-10106-6 DM 120,– £ 29.– $ 67.–
4. **Italiano** XX, 132 p. ISBN 3-598-10107-4 DM 120,– £ 29.– $ 67.–
5. **Quadrilingual Index** VIII, 168 p. and 24 microfiches ISBN 3-598-10108-2 (not available separately)

Veterinary Multilingual Thesaurus

5 Volumes ISBN 3-598-10109-0 (the set) DM 500,– £ 119.– $ 278.–

1. **Deutsch** XX, 216 p. ISBN 3-598-10110-4 DM 180,– £ 43.– $ 100.–
2. **English** XX, 244 p. ISBN 3-598-10111-2 DM 180,– £ 43.– $ 100.–
3. **Français** XX, 218 p. ISBN 3-598-10112-0 DM 180,– £ 43.– $ 100.–
4. **Italiano** XX, 204 p. ISBN 3-598-10113-9 DM 180,– £ 43.– $ 100.–
5. **Quadrilingual Index** VIII, 240 p. and 36 microfiches ISBN 3-598-10114-7 (not available separately)

3 THESAURI – 15 VOLUMES DM 920,– £ 220.– FF 2190.– $ 515.–

K·G·Saur München·New York·London·Paris
K·G·Saur Verlag KG · Postfach 71 1009 · 8000 München 71 · Tel. (089) 798901 · Telex 05212067 saurd

https://doi.org/10.5771/0943-7444-1980-1-13 https://www.inlibra.com/de/page/agb - Open Access -