

*Proceedings of the
International Study Conference on
Classification for Information Retrieval*

held at

BEATRICE WEBB HOUSE, DORKING, ENGLAND

13th-17th MAY 1957

NOTE:

- 1 - SEE PP. 91, 98 & 106 FOR EG'S
COMMENTS
- 2 - COMPLETE COPY OF THESE PROCEEDINGS
IS KEPT IN THE ISI REFERENCE LIBRARY.
- §108

ASLIB · 4 PALACE GATE · LONDON · W.8
PERGAMON PRESS · NEW YORK

1957

INSTITUTE FOR SCIENTIFIC INFORMATION
33 SOUTH SEVENTEENTH STREET
PHILADELPHIA 3, PA.

Monday 13th May: Evening session

CHAIRMAN Dr. Mayer

DR. SHERA presented his paper on 'Pattern, structure, and conceptualization in classification'.

DISCUSSION

The main theme of discussion—which is only briefly summarized here—was what may be called the 'subject-object' relation in classification. Was classification objective or subjective? If the latter, how could we hope to construct a general classification?

M. DE GROLIER stressed the objectivity of classification. Classification schemes were not simply invented: biological classifications, for example, were progressively more and more true to nature. A good classification followed the pattern of nature, as Bliss had taught. M. de Grolier disagreed with the view that it was not possible to devise a single general scheme, because each person's mind was unique. We did not classify patterns of thought, but external reality. Although a perfect, completed classification was impossible, the advance of man's knowledge brought an increasing number of subjects into the arena of 'fixed', established knowledge.

M. CORDONNIER agreed with this and said that Dr. Ranganathan had made a valuable effort to distil the principles of general classification, and had raised the science to a new level. But even his scheme was lacking in certain respects, because he had not completely put into practice his own principles. A fresh attempt at a general scheme was now necessary.

DR. HOLMSTROM warned against assuming that the same scheme of classification was applicable to all purposes—e.g. for classing objects, administrative files, books, and items of information. We must try to clarify what we meant by 'subject' at such different levels as these. There might be two aspects of classification—the classification of physical facts might be objective, but the classification of documents relating to ideas could never be independent of those using the documents.

DR. SHERA suggested that knowledge was conceptually structured in patterns determined by our mental processes. We did not discover ways of grouping, but invented them. Types of wood, say, would be grouped differently by a botanist and an architect.

MR. VICKERY pointed out that although the manner of grouping would be subjective, dependent on private purpose, the grouping would only be helpful if it corresponded to objective features of the entities grouped. It was both necessary and possible to allow for alternative methods of grouping in a general classification.

MR. FAIRTHORNE pointed out that the mathematical concept of 'congruence' allowed us to group things in various ways. We could give descriptions to things and decide whether they were congruent in particular circumstances. A complete network of relations applicable in all circumstances could not be laid down, but we could specify the pattern of the network.

MR. GARFIELD introduced a second point of discussion, that there was no conflict between classification and mechanical selection. Class numbers could be used, for example, to code punched cards. But this might lead to inefficiency, since a frequently coded concept might bear a long class number. This was a weakness of the classification, not of the machine.

M. DE GROLIER agreed that the coding requirements for punched cards were not fulfilled by traditional class numbers.

Conference on Classification

Tuesday 14th May: Evening session

CHAIRMAN Mr. Vickery

This session was a general discussion to introduce viewpoints which did not readily fit into the other sessions.

MR. GARFIELD suggested the need to explore ways of organizing information which stressed, not classification and order, but randomness and suggestiveness. There was a need to learn what was currently going on, rather than what had happened. As a device to provide such knowledge, he mentioned the dissemination of collections of contents lists of periodicals in a given field, the periodicals covered being chosen by user interest, not by their ostensible subject. He then described the 'citation index'. This was compiled by scanning the references at the ends of articles in a given group of periodicals and books. Each reference so found was listed as a heading, in any convenient manner, and under it were given references to all the articles and books which cited it. Such items might have an ostensible or primary subject far removed from that of the heading, but the fact that the two shared a common interest was made evident by the citation index. A sample survey of such an index for 5,000 patents showed that it linked together patents which were not grouped either by the Patent Office classification or Chemical Abstracts subject index. The citation index seemed to be an example of a *posteriori* organization of information, whereas organization by pre-established subject headings or classification was *a priori*.

MR. VICKERY said that the aim of classification for information retrieval was to be as little *a priori* as possible—i.e. to be based on literary warrant, but new methods of making classifications more sensitive to literary warrant were always welcome.

MISS KYLE felt that a device such as the citation index might not be very helpful in subjects which were well established, but it might be of value in rapidly developing fields of knowledge such as some social sciences. Often when a notion could not be sharply defined it might well be identified by citing previous authors who had discussed it.

DR. GROENEVELD considered that such an index would be of use mainly in speculative inquiries.

MR. FAIRTHORNE said that it could be a learning device, teaching the classifier how to classify. But it could be dangerous, because correlation in itself classified nothing.

MR. FARRADANE recalled using a technique similar to the citation index in revising an old text-book.

MR. GARFIELD mentioned another aspect of the index: under a heading referring to a worthless article would accumulate references to criticisms of it, so that those using it later would be warned.

MR. MILLS queried whether the citation index could serve as an aid to classification, since it might reflect only minor, subsidiary interconnections between references.

MR. VICKERY said that it might act as a sensitive indicator of emerging literary warrant, if we classified the citing and the cited references, and studied their interrelation. In particular, it might give an early intimation of the intermeshing of previously separate subject fields.

MISS KYLE instanced cybernetics and other new subjects difficult to classify.

MR. CLEVERDON queried whether the cost of a citation index would be justified by its use, particularly as it should be comprehensive.

MR. GARFIELD replied that it could be compiled at a fast rate and using only clerical labour.

tests on working models'). He described how a general classification might be built up by the co-operation of specialist organizations.

Later in the Conference MR. FAIRTHORNE mentioned the use of 'aspect cards' in medical diagnosis, and MR. VICKERY said it seemed possible that this was practised in Sumer nearly 3,000 years ago. He mentioned the discovery of clay tablets, each of which referred to a single symptom and listed the diseases in which it was present: perhaps the tablets were used as 'uniterm' cards in diagnosis.

Thursday 16th May: Morning session

CHAIRMAN Mr. Garfield

MR. FARRADANE introduced his paper on 'Classification and mechanical selection'.

In reply to a comment from the Chairman that the paper dealt but little with the 'gadgeteering' side of retrieval, Mr. Farradane emphasized that we had to decide what to put into a system before thinking about the machinery and the way it was to be programmed. We were dealing with information, a pattern of knowledge. Knowledge came from the mind, hence we had to understand how the mind worked. The traditional method of classification, breaking down assumed main classes, was defective: our experience developed inductively. We must build up classes from elementary ideas. But a whole was more than a simple aggregate of its parts, so it was necessary to include relational concepts between elementary ideas to bind the parts into a coherent whole. His own 'operators' were an attempt to state explicitly the basic relations derived from psychological evidence. They were used to build 'analets' which could be fed directly into a machine or used to construct classifications of the old type as required.

DISCUSSION

The discussion ranged over a wide field. Much of it was taken up with the problem of 'relations' in information retrieval, but some aspects of machine selection were also considered.

1. *Relations in information retrieval*

MISS KYLE began the discussion by suggesting that Mr. Farradane might be trying to make a more logical language for accurate communication rather than just for retrieval purposes: he was desiderating a greater precision for classification than was used even by authors of the material classified. On the question of his notation for indicating different types of relationship, she asked whether a simple numbered listing of his nine relational operators would serve.

MR. FARRADANE said that they had to be used with care—the operators were directional and they expressed psychological relations not capable of comprehensive definition in ordinary language.

MR. GARFIELD gave examples of specific directional relations already used in mechanical selection systems.

MR. FARRADANE replied that to use a large number of specific relations, dependent on particular 'isolates', peculiar to special situations, was unsatisfactory. He reduced the multiplicity of relations to a small set of general operators.

DR. MAYER later suggested that the relational analyses of both Mr. Farradane and Dr. Ranganathan were biased by their use of the English language.

Summary of the Discussions

The statement that the operators were not easily expressed in ordinary language was queried by several speakers.

MR. WRIGHT said that we started from documents written in ordinary language and so had to use it.

MRS. BROWNSON did not see how we could get away from language in constructing a classification.

MR. WELLS and MISS KYLE suggested that Mr. Farradane was classifying a large number of 'ordinary' particular relations under nine heads—the operators.

This led DR. HOLMSTROM to suggest that it was preferable to list a large number of particular relations.

M. DE GROLIER said that many relations were needed, and he queried whether relations (morphemes) and substantives (semantemes) should be distinguished from one another.

DR. RANGANATHAN considered that substantives were less abstract than relations, and that the use of the latter required more careful control. He urged the need for the comparative study of many fields of knowledge to elucidate the relations needed and to classify them.

MRS. BROWNSON mentioned the project of Miss T. M. Williams to analyse a set of titles of scientific and technical papers to determine what conceptual distinctions and relations were at work and to systematize these into a system suitable for searching. The titles to be used were those of the 1,100 papers submitted to the International Conference on the Peaceful Uses of Atomic Energy (Geneva, August 1955). Mrs. Brownson suggested that if other such analyses were to be made, using different techniques, it might be helpful to use the same titles so that comparison of results would be facilitated.

MR. WILDHACK said that in a small collection of information sufficient discrimination in selection was achieved by the joint use of terms, and only as the size increased was it necessary to specify relations between terms. We needed to establish at what point this occurred.

MR. FAIRTHORNE also stressed this. There was no need to build into a system more discrimination than was required to select a given document. In relatively small collections (say, 5,000 to 50,000 items) relations were not needed. But since collections constantly grew, 'openness' had to be maintained, and it should be possible to insert relations when required without upsetting the system.

2. Machine systems

The discussion of mechanical systems of information retrieval remained fairly general.

MR. VICKERY said that he felt the need of answers to such questions as: What conditions did machine systems impose on classification schemes used in them? What problems would be encountered in the application of schemes to mechanical selection—e.g. in coding?

MR. FAIRTHORNE pointed out that all the machine did was to reduce the need for human clerical labour. For it to do this, men had to instruct the machine in detail. We had to distinguish between real thinking and mechanical (clerical) work, and to decide what we could delegate to the machine.

MR. FOSKETT said that we needed an ordered structure of knowledge in any retrieval system, and a notation to apply it in the system. A simple manual system sufficed in certain circumstances. At what point did the use of a machine become economical?

MR. GARFIELD outlined four areas in which machines could be used: the mechanical analysis (indexing or classifying) of actual documents, including mechanical reading of the text; machine indexing—i.e. an indexing system which operated mechanically; mechanical aids to searching an indexing system; and machine listing—the preparation of conventional indexes by machines.

MR. FAIRTHORNE added a fifth use, 'learning'.

M. DE GROLIER raised several problems concerning coding for machine selection. Was it possible to use a single code for the three purposes of (a) arranging documents, (b) selecting references, and (c) analysing the contents of documents? He thought probably not. The type of coding was clearly affected by the machine used, but what were the relative advantages and disadvantages of, for example, (a) completely random code symbols, (b) 'semi-random' symbols derived from a natural language, or (c) systematic, hierarchical code symbols? He himself was doubtful of the value of self-demarkating codes and of superimposed coding. The most 'hospitable' machines seemed to be those at either end of the scale of complexity—on the one hand, simple systems using 'aspect cards', and on the other, large electronic computers. The choice of machine would depend on the size of the collection and the types of use.

M. CORDONNIER presented a cross-classification of selection systems. Four characteristics of division were used:

A: By the form of the record

- 1: 'memory box' (as in a computer)
- 2: individual card for document (e.g. microfiche)
- 3: aspect card (Selecto, Peek-a-boo)

B: By the material recorded

- 1: the whole document
- 2: an abstract
- 3: a reference
- 4: a call number

C: By the physical form of coding

- 1: by inscriptions on a tab
- 2: by slots
- 3: by perforations
- 4: photographic
- 5: magnetic

D: By the symbolization of relations

- 1: indeterminate
- 2: by position
- 3: by grouping of code elements
- 4: by shape of code elements

There were many possible combinations of these factors. Each combination should be assessed economically. Should we aim at maximum simplicity in each characteristic, or at complexity?