

C O C T A

Committee on Conceptual and Terminological Analysis
International Political Science Association

9th Congress, Montreal, 1973
Research Committee 1,
(Session 3, August 24th, Room 621, 3-6 pm)

TOWARD A CONCEPT INVENTORY

suggestions for a computerized
procedure

Anthony Judge

This paper is the revised and abridged version of Working Paper No. 3 prepared for the Committee on Conceptual and Terminological Analysis [COCTA] of the International Political Science Association. The paper constituted a much-expanded version of a set of notes which were discussed in relation to proposals in COCTA Working Paper No. 1 (prepared by Fred Riggs, Secretary of COCTA), and in the COCTA Manifesto (prepared by Giovanni Sartori, Chairman of COCTA, and Fred Riggs) at a meeting sponsored by the International Studies Association and held, on the invitation of the Rockefeller Foundation, at the Villa Serbelloni (Bellagio, Italy, 1-5 September, 1971).-

Union of International Associations
1 rue aux Laines, Brussels, Belgium
May 1973

CONTENTS

Introduction/Summary	3
Project Objectives	
Concept Inventory Phase	4
Concept Classification Phase	4
Term Allocation Phase	5
Organization of Project	5
Classification and Modelling	9
Entities, Relationships and Models	
Types of Entity Included	12
Types of Relationship Included	13
Types of Model	15
Data to be included on each Entity	21
Limitation of Scope and Sources of Concepts	30
Concept Notation in Documents	32
Representation of Concept Networks using Graph Theory	33
Use of Interactive Graphic Display Techniques	40
Implication of Computer Argumentation of Intellect	44
Relationship to Artificial Intelligence Projects	* 45
Other Initiatives	
Concept coding schemes	47
ADMINS system	48
Citation indexing	48
Subject classification schemes	51
Concept dictionaries	53
World problems identification	54
Language and Translation Problems	55
Documentation or Knowledge?	
Documentation problem	58
Knowledge representation	61
Summary of Advantages of this Proposal	65

"The possibilities open to thinking are the possibilities of recognizing relationships and the discovery of techniques of operating with relationships on the mental or intellectual plane, such as will in turn lead to ever wider and more penetratingly significant systems of relationships."

(B.L. Whorf. Language, Thought and Reality)

INTRODUCTION

This paper addresses itself to the practical problems of developing a means of filing concepts and other theoretical constructs in a data bank. Such concepts would be filed as entities having a distinct meaning and not in terms of the word by which they happen to be represented in a particular school of thought. The reason for this approach is that many of the words on which most reliance is placed in the social sciences (e.g. "group", "class", "power", or "structure"] have acquired a multiplicity of overlapping meanings (+).

The concept file so created would be used to generate lists, to facilitate classification and interrelation of concepts to produce concept thesauri, and, finally, to facilitate the allocation of "authoritative" terms to permit the production of terminological thesauri.

The object of this project would be to ensure that any qualified person – with a few safeguards – would be free to register entities in the file which would then become available for secondary analysis at any interested research centre.

One form such analysis might take would be the construction and comparison of various models or classification schemes for theoretical entities. At a tertiary level, efforts could be made to link such entities with each other, cutting across the boundaries of disciplines, ideologies, epistemological approaches, paradigms or problems. This activity would provide new alternative means of approaching the entities held on the file but would not affect their use for more restricted purposes.

In this paper particular attention has been paid to some of the techniques available to analyze complex entity networks or structures. Because of this complexity and the problems of comprehending it, the use of interactive computer graphics has been examined as a powerful means of simplifying the task and making the project more widely significant.

(+) Fred W. Riggs, "Concepts, Words and Terminology". Honolulu, University of Hawaii, Social Sciences Research Institute, 1971, 66 p. (Committee on Conceptual and Terminological Analysis, Working Paper No. 1].

Giovanni Sartori, "Concept misinformation in comparative politics". American Political Science Review, 64, December 1970, 4, p. 1033-1053.

PROJECT OBJECTIVES

A project to handle, structure, and analyze theoretical constructs is proposed which would be operated as three distinct phases:

- a concept-filing or registration phase leading to the creation and maintenance of concept inventories
- a concept classification phase, leading to the production of concept thesauri
- a term-allocation phase leading to the production of standardized terminological thesauri

A translation phase, to make the project more widely relevant, would run in parallel with the above three. Each succeeding phase builds on the previous one, but need not necessarily follow it immediately in time for the project as a whole to be of value.

1. Concept Inventory Phase

A computer-based concept registration or tagging system should be set up which would allocate sequence numbers to concepts on a continuing basis. The criteria for concept registration should be kept to a minimum to ensure that the system remains "open" to a wide variety of users and contributors.

This approach permits rapid inclusion and organization of the data and rapid production of updated concept lists. These would facilitate the scrutiny of the data in later phases and in terms of the perceptions of different need groups.

2. Concept Classification Phase

#

Evaluation, classification and identification of concept interrelationships would be made independently by a limited number of contributing groups, possibly associated organizationally with the international academic bodies. These groups would be primarily concerned with allocating codes to be fed back to the computer system so that ordered and refined concept thesauri could be produced to reflect the perceptions and needs of the contributing groups. An important aspect of this coding function by groups would be the rejection of those concepts registered which are considered to be of little value to the group's perspective.

From the computer data handling point of view, each contributing group would be building, refining, and maintaining its own "model". Each such model would be handled as an independent optional qualifier on the sequentially-ordered concept list.

From the point of view of any such group, the computer system would be viewed as holding the concepts in which it is interested in the order of its own preferred classification scheme.

There would of course be the opportunity at any time to look at the same concept list through the classification scheme of any other contributing school of thought. Concepts would be identified by their sequential number plus a number which would identify the model employed.

3. Term Allocation Phase

At a later stage users of one model might find it useful to produce an "authoritative" list of terms to be used for those concepts of interest to them. This could also be incorporated into the computer system. Such terms could then be used to produce standard terminological thesauri for the users of one model.

ORGANIZATION OF PROJECT

The success of a project of this type would be dependent upon the extent to which any central organization can be avoided in favour of a process of catalysis. There is too much to be done to run the risk of the usual jurisdictional, behavioural, and personality problems associated with a centralized organization. Such problems rapidly alienate potential support. The problem is therefore to bring into existence a decentralized network of groups working on different aspects of the project, but able to exchange the results of their activities without difficulty.

(It is important to remember that it is probably impossible to "organize" a whole area of knowledge because the latter is well subdivided into territories and "stamping grounds" whose incumbents are reasonably content with the current situation. It may, however, be possible to offer them a reasonably neutral device by which they can each facilitate and order their own particular approach, and, as a by-product, see more clearly its relationship to that in other "neighbouring" territories. Having by this means obtained a decentralized picture of the current situation, it is then possible, in a totally distinct process, to lobby the incumbents into participating to some degree in inter-territory efforts at organizing areas of knowledge whilst guaranteeing safeguards for the protection of their "sovereignty".)

A. Launching Phases

A number of Phases can be envisaged, some of which could overlap.

1. Investigation. During this Phase the project would be investigated in detail by circulating proposals among appropriate specialists. The main object would be to ensure that the proposal is oriented in the right direction, and that funds for pilot projects are obtained. This Phase may be considered to be underway already, through the actions of the COCTA committee.
2. Pilot Projects. During this Phase efforts would be devoted to the following areas:
 - a] computer program development and file organization.
 - b] operational and logical problems of classification with models in a few test areas.
 - c] computer simulation of file movement and modelling activity in a decentralized, minimum-organization environment.

It would be particularly valuable to gain some insight into the behavioural problems of rivalry and suspicion between model building groups, and efforts to "take over" the system.
 - d] computer simulation of different strategies to keep the system "open" to theoretical formulations from as wide a range of sources as possible whilst trying to minimize the inclusion and retention of formulations of dubious value.

3. Agreement on Standard Formats. On the basis of the previous Phase, standard formats for filing new formulations and for holding them on magnetic media would be agreed. Since this is a new type of project, it should not encounter the apparently insurmountable difficulties of those concerned with organizing the computerized exchange of bibliographical information.
4. Production of Standard Software. Once agreement has been reached, a standard software computer program can be made available to all those bodies which wish to initiate some concept modelling activity, to act as a central filing point for their particular constituency. It is possible that initially only one body will be active, possibly as an extension of the pilot project stage.
5. Filing Procedure. Once a standard filing or registration form is developed, there should be no difficulty for any group in receiving and filing identified concepts. This can of course be done by mail.

By filing is meant the purely administrative activity of preparing the forms for the computer. There should be a minimum of judgemental effort at this stage, and none with respect to the theoretical problems of the subsequent modelling activity. The object is to get the incoming information into a form which facilitates the activities of the members of the modelling bodies.

The area of difficulty which does require examination is that of how to decide who should not be permitted to submit concepts for filing into the common data base. This point is considered below.

B. Periodic Operations

1. Lists of Formulations. Periodically the sequence of identified concepts held on magnetic tape should be scanned to produce lists for circulation to the modelling bodies and, if required, their members. Two types of lists can be envisaged.
 - a) lists of newly-registered concepts which must be scanned by each modelling body to see whether they are in any way relevant to its concerns
 - b) lists of the complete sequence of concepts for newly formed modelling bodies wishing to re-examine all possible formulations and interrelate them in their own way.
2. Modelling or classification. The lists derived from the previous operation can be examined by the modelling bodies in committee or distributed by post to their members. From these [postal] deliberations should emerge a collective opinion on the place within the classification scheme, of each identified concept reviewed. If necessary, a "provisional" view can be formulated by the use of appropriate coding. In fact this might be a most useful way of submitting a committee's view for wider consideration. Different degrees of "definitiveness" could thus be envisaged.
3. Feedback of Model Information. The details of the place of the concept within a particular model would be indicated on a standard form which could be returned by post for keypunching and incorporation. A modification of this approach would be to permit individual committee members to each return forms for any new entity under consideration. In this way all the alternatives would be incorporated into

the model with some "provisional" code so that each member could see the proposals of the others, and their implications. In some cases, this could even be operated as a means of postal voting on the treatment of controversial concepts. The administrative load of the committee is in this way largely computerized.

4. Input of Model Information. The forms from each modelling body would be handled at the central registry point [for that constituency), keypunched and fed onto the magnetic tape file. Key-punching errors would be corrected there as far as possible.
5. Production of Model Amendment Lists. Whenever required, the concepts incorporated into a given model would be selected and sorted into the thesaurus-type structure appropriate to the model and listed for distribution back to the members of the modelling body. This gives members an updated model with all the concepts coded to different levels of "provisionality".

Members can then reconsider their views and proceed from Operation 2 above or, alternatively, for those formulations which have been classified to the agreement of all concerned, the term allocation operation may be initiated.

6. Allocation of Model Terms. Working from the concepts structured into a thesaurus-type order, members can allocate terms to each entity in English and whatever other languages are considered necessary. Again, there is no reason why "provisional" coding should not be used to cover various working cycles of term allocation.
7. Feedback of Term Information. As with model information, the alphanumeric terms allocated to each concept can be indicated on a standard form which could be returned by post for keypunching and incorporation onto magnetic tape.
8. Input of Term Information. The forms from each body allocating terms within a model would be handled at a central registry point, as with the model information itself.
9. Production of Term Lists. Whenever required, the concepts incorporated into the model would be selected and sorted into term lists, either in alphanumeric order or in terms of a thesaurus-type structure. This gives members an updated model expressed in terms coded to different levels of "provisionality".

Members can then reconsider their views and proceed from Operation 6 above.

It is clear that the above operations permit a quite extensive degree of "de-committeefication". Members of a modelling body can individually register their views and preferences by post on each concept in the model and in their own time. The resulting lists are circulated and amended to firm up progressively the consensus on each point until final agreement can be reached. Alternatively, if this is a final difference of opinion, then this can be registered as such. Actual discussion need only take place when the accumulation of cases (which cannot be handled by correspondence and a "modelling bulletin" mechanism] merits such contact.

Subsequent Phases

A number of Phases can be envisaged which follow on from those detailed in "A" above. They do not, however, modify the basic operations noted in "D".

1. File Movement. One of the disadvantages of isolated registration points is that concepts common to two or more constituencies will not necessarily be juxtapositioned. In particular, unless each such point is allocated a block of sequence numbers, there are liable to be overlapping sequence numbering systems which would jeopardize the whole project. •

One means of avoiding this, aside from allocating blocks of numbers to each registering point, is to circulate copies of the files between registration points. (Either the tapes themselves could be moved, or data links could be used.) This might be considered a standard procedure by which duplicates in all newly-coded concepts could be located and grouped together for consideration by each of the interested modelling bodies, prior to arriving at a "final" decision.

The circulation of such information can be made very rapid. A courier file can be circulated between the registration points for a particular discipline. Information is copied onto and off each such sub-specialty file. At one point in its movement, such an intra-discipline file could interact with an inter-discipline file (e.g., for disciplines in the same group] to permit a similar two-way transfer to take place. Similarly a higher level courier file moving between groups of disciplines could permit further exchange.

In this way cross-disciplinary confusion could be avoided. Clearly refinements are possible by using mission-oriented files or geographical area files. The system is very flexible. It could even be made to interact with "classified" files by using security, subject matter and evaluation filters to govern the interaction.

The key feature is that it does not require more than a bare minimum of overall organization or funding. It can be extended very loosely in response to the initiative of any highly specialized discipline. Registration points are created wherever (in terms of subject, jurisdiction or geographical level] there is sufficient common interest - i.e., motivation plus resources. This gets around the current situation in which vain attempts are made to get significant funding for general multi-purpose projects, particularly via any international program.

If cross-jurisdictional problems arise in particular areas, all the administrative work there may be delegated under contract to some party judged to be impartial and uninvolved -- a commercial computer service bureau, a university, a government agency; or a user cooperative point might be organized.

The costs involved at each collecting point are

- (a) conversion of information and queries to machine-readable form
- (b) processing and output relevant to immediate user contacts
- (c] transport costs of the courier file to the next collecting point.

The funds are expended locally in a manner which can be immediately justified and yet this results in making available current information from points very conceptually distant within the system.

Accredited Sources

It is clearly an advantage to allocate responsibility for modelling group activity in a particular domain to the appropriate international professional organization.

The difficulty arises in determining which sources of information should be recognized by such modelling groups. In the earlier phases when the group is working through the standard texts, few problems should arise. But once a model is available for inspection, problems will arise in determining whose suggestions for additions or amendments should be accepted. Within a well-defined profession this difficulty may be avoided by recognizing only accredited members of the profession. The right to submit amendments then becomes a right accorded by the profession. This procedure will undoubtedly lead to conflict when areas common to a number of disciplines are considered (e.g., the social sciences, in general), unless each discipline is restricted to its own model.

A distinction should also be made between the right to file an entity and the right to suggest amendments to the model. There is some advantage in giving wider access for filing, but limiting the "retention period" of the entity filed according to the professional standing of the filer.

A later development could be the possibility of retaining entities only if a supporting "vote" was registered by an appropriate number of appropriately accredited persons. The degree of support would be a "real time" measure of the degree of significance to the discipline of a given theoretical formulation.

Whatever procedure is adopted it is essential, for the vitality and general relevance of the project, that a wide range of people and organizations should be in a position to add entities to the file -- given a few simple safeguards. In this way the interests of every relevant discipline, school of thought, problem area, "approach" or paradigm should be protected. The system would therefore be "open" to social scientists writing in any language or taking any epistemological or ideological position.

CLASSIFICATION AND MODELLING.

1. Nature of Classification

There is a considerable terminological variation in the scientific literature that characterizes the use of the term "classification". Dalenius and Frank, after making this observation CD define the term as follows.

"Consider a collective of objects of some kind and a set of mutually disjoint classes. Every object belongs to one, and only one, of these classes. By classification we will denote the act of assigning the objects into these classes.

CD T.E. Dalenius and O. Frank, "Control of classification". Review of the International Statistical Institute, 36,3,1968, 279-295 (includes formal description of classification and introduces various parameters useful for control purposes].

In taxonomy, classification indicates the act of creating classes according to some principle, the term "identification" is used for classification as used in this paper. By the same token, the term "coding" is rather ambiguous. We refrain from its use here, but mention that classification as used in this paper is referred to as coding in **the** literature dealing with e.g. population censuses."

This definition, whilst appearing to be inclusive, in fact only covers **one** type of classification, namely hierarchic classification where classes are mutually disjoint. Classification of theoretical formulations is one area in which classes may or may not be mutually disjoint.

J.H. Shera has made an excellent general assessment of the problems of general library classification in an article of his, originally published in 1953 and reprinted in his book, Libraries and the Organization of Knowledge. He concludes that the hierarchical form in itself is not a sufficient basis for the classification of knowledge and that what is required is a directed graph, or non-hierarchic representation.

The relationship between hierarchic and non-hierarchic classification schemes has been the subject of considerable work by Jardine and Sibson (1). They are particularly interested in the stability of the classification produced by a given method as the amount of information [number of attributes] is increased for the entities being classified. They are looking for measures of distortion introduced by the imposition of a given classification scheme.

This work makes it clear that the process of classification can introduce distortion and that this can be avoided by using a directed graph representation. In this project the distinction is made between the filing process, the classification process, and the term allocation process.

It is useful to think of the first stage of the classification process as one of "relationship indication", in which the relationships of a given theoretical entity with other entities are inserted. This results in a [directed graph] network of entities which can be searched by computer, particularly to detect clusters with certain properties. This stage corresponds to the determination of similarity or dissimilarity between entities.

In a second stage, the above network can be distorted so that its elements can be fitted into a chosen set of classes with a certain relationship to one another. This is "classification" as opposed to the previous phase which inserts relationships irrespective of any class boundaries. It is convenient to call this activity "modelling". Clearly the modelling activity is a valuable preliminary to "classification". It is particularly valuable in that once completed, different systems of classification can be compared using the entities inter-related by the model, i.e. different degrees of distortion can be imposed upon the network of entities according to the immediate needs of the user. It may be useful to think of modelling in this context as a long-term multi-person activity, whereas a given classification can be selected from the modelled entities in terms of short-term, need-oriented considerations which permit certain

[1] N. Jardine and R. Sibson. Mathematical taxonomy. London, Wiley, 1971.

relationships in the network to be considered as "irrelevant" -- permitting the isolation of simple, possibly hierarchic, classification schemes. In some cases, it may however be preferred not to distinguish modelling from classification and to blur the two operations into one another.

2. Filing and Classification

In the case of document indexing application, no distinction is made between filing and classification. Because of this, the administrative problem of filing and the qualified expert problem of classification combine to create severe problems.

The UNISIST(1) Study noted that little progress can yet be reported in the way of indexing-at-source and that a serious limiting factor to any form of cooperative indexing is the range of acceptability of the proposed indexes. Even the all-embracing and widely used U.D.C. has adversaries. The Study also noted that it is unlikely that the concept of a universal scheme will ever make any practical sense in the realm of deep content analysis (p.46). . . The reasons are the observed differences in the semantic basis of indexing languages which are the consequence of well-founded differences in outlook and interests on the part of a highly-diversified community of users.

All that can be looked for, according to the UNISIST Study, "is the existence of semantic relations between the different lexical sets (be they called classifications, lists of disciplines, thesauri, automatic dictionaries for converting natural language into information language, etc). The study of these relations is the subject of ongoing research on the "compatibility" of indexing vocabularies the subject is now receiving much attention as an essential part of projects aimed at establishing world-wide interconnections between information systems." [p.46]

It would appear from this, that the distinction between the impracticalities of classification and the practicalities of "relationship identification" (i.e. modelling) is becoming established. But the filing or administrative aspect of "entity capture" is now blurred into the modelling phase. There is as yet no suggestion that work on "compatibility" would be considerably facilitated if similar filing techniques were used prior to the activity at the modelling level at which the "well-founded" theoretical differences arise. Standardization is possible, but at a lower level consistent with user requirements. Until this is realized the relationship between lexical sets cannot be handled systematically by computer methods.

3. Advantages of Numerical Filing System

The three major advantages of a sequential, non-significant numbering system for entities are -- facilitation of administrative activity by removing the burden of requiring that the file number receive the "imprimature" of an overloaded qualified expert

-- preparation of the basis for a proper semantic analysis by avoiding "the difficulty encountered in manipulating semantic

(1) UNESCO, UNISIST; study report on the feasibility of a world science information system. Paris, Unesco, 1971.

reality without the assistance of a corresponding concrete reality" (1) and permitting "semantic facts to be treated independently of their formal (linguistic] supports"(1).

- admission of "artificial" theoretical entities (new concepts, groupings of other concepts) for which no simple term exists or for which a questionable neologism would have to be invented. This is difficult in the case of term oriented systems.

ENTITIES, RELATIONSHIPS AND MODELS 1.

Types of Entity Included

There is a very varied terminology currently in use to characterize theoretical products. Gunnar Sjöblom notes the use of conceptual (analytical, theoretical) frameworks, analytical schemes, paradigms, orientations, frameworks for inquiry, theory sketches, pre-theories, etc.(2) The same is true for the components of the scientific process: problems, observations, empirical generalizations, models, derived propositions, hypotheses, theories, etc. It is unlikely that any immediate agreement could be achieved on a standard terminology, even if this was in fact beneficial.

Each of the conceptual constructs represented by the above terms may be treated as an "entity" which could be incorporated into a computer file. Once incorporated, efforts could be made to attach an appropriate distinguishing code to them within the framework of a given model. It is highly probable, for example, that under different models the same entity may be coded differently, or alternatively that distinctions important within one model will be insignificant in another (e.g., theory and model; hypothesis and proposition).

As a summary, the above entities are numbered below to facilitate discussion on possible groups of entities:

A. Concepts

B. Meta-concepts

- | | |
|-----------------|--------------------------|
| 1. theories | 5. analyses |
| 2. propositions | 6. conceptual frameworks |
| 3. hypotheses | 7. analytical schemes |
| 4. models | 8. theory sketches |

C. General

1. paradigms
2. viewpoints
3. schools of thought

(1) A. Martinet. "Arbitraire linguistique et double articulation." Cahiers Fernand de Saussure, 15, 1957, 107 (cited by Georges Mounin, Les problèmes théoretiques de la traduction, Paris, Gallimard, 1963, p.122-123).

(2) G. Sjöblom. Theoretical testing of approaches in political science (Paper presented at a conference of the International Studies Association, Dellagio, 1971).

D. Assumptions

1. assumptions
2. criteria
3. values

E. Methods

F. Problems

1. substantive
2. methodological
3. problem formulation

G. Hierarchies

1. taxonomy
2. typology
3. classification

H. Operationalization

1. indicators
2. indexes

I. Data

1. bodies of data
2. interpretations of data
3. observations

J. Social

1. organizations

There is some advantage in a two-level coding here, because it might be possible to arrive more easily at agreement on the more general level coding, even if there are differences between models on the coding within that level. There is of course the possibility that within a particular model the grouping would be done differently, in which case the coding scheme would be peculiar to that model.

2. Types of Relationship Included

It is not the intention of this project to set up a single rigid classification of permissible relationships between entities. Just as no effort was made to limit the types of entities that could be handled (see above), it should not be necessary to make the futile attempt to resolve the intellectual problem of how many types of relationship are significant. That the attempt would be futile on the part of any one group is shown by Eric de Grolier's excellent chapters on the expression of relationships in generalized and specialized coding systems, in natural languages, and in

experimental languages CD. He concludes, in his UNESCO/FID supported study, that it proved impossible to produce a systematization that was "sufficiently satisfactory to warrant even preliminary publication".

This conclusion should not however lead to a decision to adopt some hypothetical "best existing scheme" or to the formulation of a new scheme. It should be recognized that the project should be capable of handling as many different schemes as possible. In fact the evolution of Knowledge is partly represented by attempts to produce new schemes of relationship and categorization.

Without recommending any particular scheme, it is useful to attempt to list out some of the relationships to give an idea of the variety that has been envisaged. De Grolier suggested a clarification of the sign ":" in the UDC (rejected by the FID Central Committee on Classification for the UDC) which covered the following relationships:

1.1 Appurtenance (belonging]

- 11 Inclusion, implication
- 12 Parts, organs
- 13 Components, constituents . .
- 14 Properties, attributes
- 141 " " physical
- 142 " " chemical
- 143 " " biological
- 15 Aptitudes, predispositions

1.2 Process

- 21 Action: acting on (subject], affected by (object]
- 211 Favourable (stimulation; increase]
- 212 Unfavourable
- 2121 Delay
- 2122 Inhibition
- 2123 Destruction
- 21 Interaction
- 211 Favourable (symbiosis)
- 212 Unfavourable (antagonism, competition)
- 22 Operation, means used: process (subject], product, result (object]

1.3 Dependence

- 3 Causality, origin, etc.
- 31 Causality; cause (subject], effect (object)
- 32 Origin: . originating (subject), arising from (object)
- 33 Conditioning, requirement: conditioning (subject), conditioned (object)
- 3 Interdependence
- 31 Correlation
- 32 Association
- 33 Combination, synthesis

(1) Eric de Grolier. A Study of General Categories applicable to Classification and Coding in Documentation, Paris, UNESCO, 19B3, p. 17-60, 61-142, 143-158

1.4 Orientation

- 41 Aspect, particular case
- 42 Application
- 43 Use

1.5 Comparison

- 51 Resemblance, likeness, similarity
- 511 Analogy
- 512 Equality, identity
- 52 Dissimilarity, unlikeness
- 521 Difference
- 522 Opposition (of character)

Other typologies of relationships have been formulated by Gardin, Farradane, Perry and Kent, Juilland. Each uses very different and, at least superficially, unrelated categories. The different models envisaged in the next section encompass compositional, behavioural, didactic, historical, cybernetic and problem-oriented relationships.

3. Types of Model

It is important to keep in mind the many possible uses of the proposed computer-based filing system. Concentration on one set of uses may not necessarily keep the system alive either in terms of funding or value to current research activity. Multiple demands on it would ensure multiplicity of fund sources and many bodies willing to feed in entities and assist in different aspects of the coding.

The following types of model are an illustration of the possible lines of development. The list does not pretend to be exclusive so that other kinds of model could be included. An attempt has been made to group the models into types which in some cases might usefully be treated on the same occasion by the responsible modelling group.

It is important to note that the models are not only simple hierarchies but can also be networks of relationships in cases where categories overlap or one entity can be a component of several other entities.

Group 1: Current Structures

This is a poor title but refers to all the current and new structures and relationships as made up of:

1.1 Compositional Models

These models would be primarily concerned with the manner in which entities are nested within one another to form hierarchies. Six types of relationship are possible here in three sets of two.

- a) Meta-level: reference numbers of all entities of which this entity is a component.

(This relationship could be split into two sub-types as the computer-level data formats for other types of model require such a split.)

Examples are: theories in which this concept is used, general class of concepts to which this concept belongs, general problems of which this problem is a part, organizations of which this organizational unit is a member.

- b) Sub-level: reference numbers of all entities which are components of this entity. [This relationship could be split into two sub-types for the same reasons as above.)

Examples are: concepts used in this theory, concepts which belong to this class of concepts, properties or attributes of this concept, sub-problems of this problem, organizational units which are members of this organization, etc.

- c) Associated reference numbers of all relevant entries which have a horizontal relationship to this entity.

See - also entities, namely those which should also be borne in mind when considering this entity.

Examples are: cases of insufficient terminological precision.

- Use - instead entities, namely those which should be substituted for this entity.

Examples are: cases where an entity is outmoded for that model.

An interesting map of relationships between conceptual entities is given in figure 1. This shows the interlocking and meeting of concepts associated with measurement of simple physical phenomena.

1.2 Behavioural Models

At the same time that the modelling activity is undertaken on the compositional relationship in 1a, it should be useful to consider some non-compositional relationships to other entities. In other words, the effects of the presence of one conceptual entity on another in the "ecosystem of ideas" (1). By this is meant concepts which are indirectly undermined or strengthened by the validity of this concept, organizations whose monopoly is weakened by the presence of this organization.

Group 2: Contextual Structures

Again -this is a poor title but refers to the historical and comprehensional relationships which constitute a context for the Group 1 current situation, and would be used in learning about the Group 1 situation.

2.1 Educational Models

These models would be produced by those modelling groups primarily concerned with education and making more sophisticated concepts comprehensible.

(1) A term suggested by Geoffrey Vickers. Value Systems and the Social Process. London, Pelican, 1971.

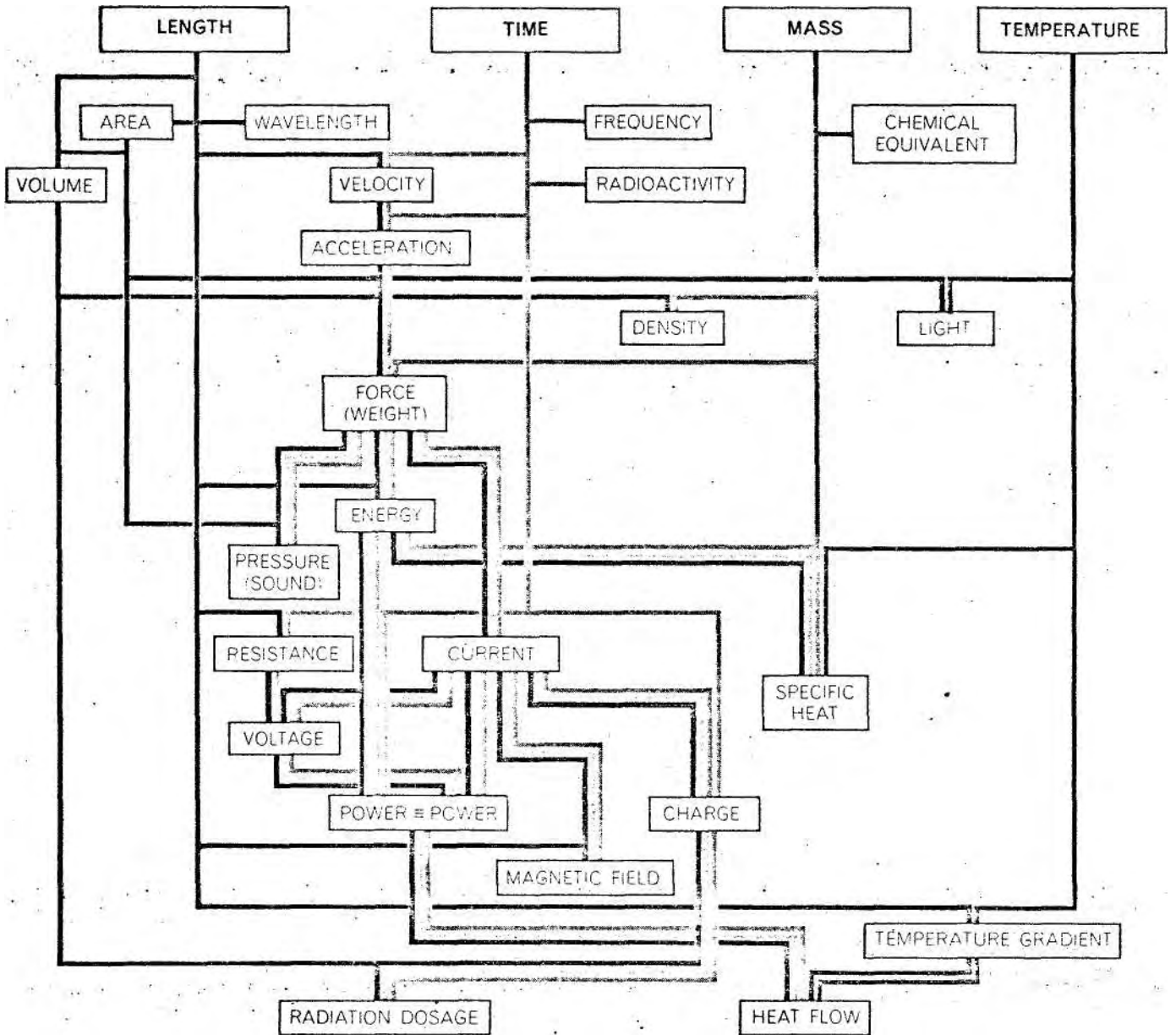


Figure 1
INTERRELATION OF MEASURING STANDARDS provides a capsule course in elementary physics. Thus such quantities as velocity and acceleration are functions of distance and time. Force,

energy and pressure involve not only distance and time but also a third quantity: mass. Many important quantities such as inductance, capacity and viscosity have been omitted for simplicity.

(reproduced from Scientific American,
June 1968, p. 62)

2.2 Historical Models

These models would be produced by those modelling groups interested either in historical research on the history of ideas or in providing an historical framework to assist education. It is probable that the educational and historical models should be considered together, which is why they have been grouped.

Group 3: Real World Systemic Relationships

The previous groups of models deal with the relationship between conceptual entities in anthropocentric terms or within the logic of particular disciplines, It is also useful to consider the systemic effects of real world entities on one another. This produces another pattern of relationships between the entities registered.

The best example of this distinction is the inter-disciplinary nature of environmental problems, when for example, it is the real world interaction of chemicals in food chains which cause egg shells to become thin --leading to high chick mortality rate of some bird species. For a social example, the relationship shown between the entities, represented by boxes in figure 2, give a schematic representation of the factors binding a Canadian Indian to a pattern of problems.

Group 4: Term-oriented Models

In some cases where classification is rudimentary or non-existent, the emphasis is placed immediately on the terms. This is the case when:

official terms are used and the definitions are conventional or undefined as in many library or descriptor lists. The entity is defined by the term.

a particular official definition exists for a particular term as in official dictionaries (e.g. the Larousse Litre as reflecting the decisions of the Académie Française)

terms are related in a thesaurus without definitions (e.g. as in Roget's Thesaurus]. Such thesauri may have many levels of classification.

There is no reason why each such set of terms should not be treated as a model as in the other groups. Where appropriate, the classification code position would be omitted and only the term positions used.

Group 5: Administrative Models

The assumption made in discussing the earlier groups of models was that the model was in some way a definitive structure on which new work would build. It is however possible to use the model building code to facilitate the administrative work on the definitive model.

Group 6: Mission-oriented Models

An assumption made in earlier groups is that the modelling bodies would all be discipline-oriented. There is however no reason why mission-oriented models should not be used where appropriate (e.g. in connection with development, environmental problems, etc.]

PRESENTACION ESQUEMATICA DEL PROBLEMA INDIGENA EN CANADA

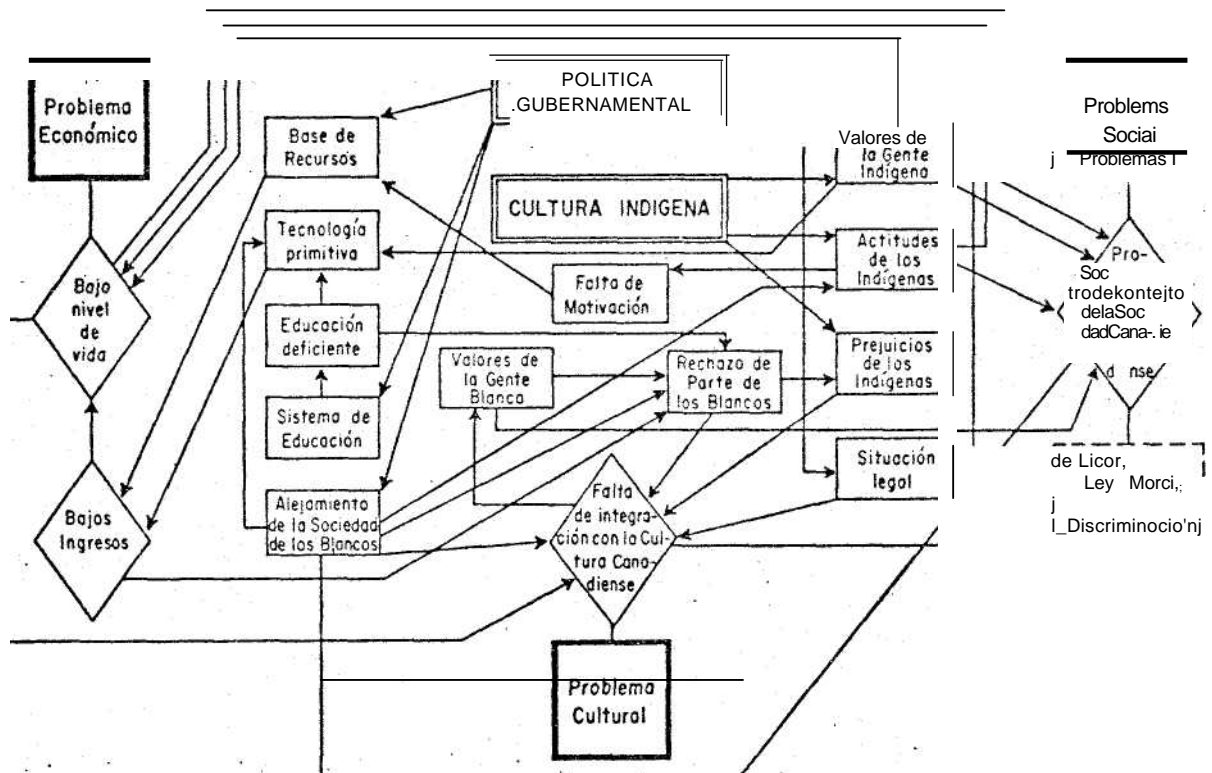


Figure 2

Illustration of cybernetic interactions

Group 7: Interdisciplinary Models

Clearly it is most important to avoid a "babble of models". A second level operation of model reconciliation to form a set of interdisciplinary or inter-model models could therefore be instituted when required.

These could either (i) be constructed (automatically by computer) from all the entities common to the models from which it is desired to produce an inter-disciplinary model, or (ii) be constructed by selection based on judgement of the best from each.

Group 8: Future-oriented Models

A final assumption made in dealing with the earlier groups was that only the current or historical situations would be modelled. There is however no reason why speculative models should not be produced showing the relationships between entities at different points in the future. The modelling activity might then in some ways represent the Delphi method of forecasting.

Group 9: Personal Models

Perhaps a long term ideal is for a person to be able to "look at" (or interfere) with the basic list of entities in terms of his own model which is his personal "thought file". Each new idea he gets could be usefully reflected in the structure of this file.

Group 10: Sub-models

In some cases a particular sub-branch of knowledge may be fragmented by re-interpretation, reconceptualization and redefinition of the same entities. It is then appropriate to use a "sub-modelling" strategy. In other words, instead of requiring "dissident" groups to conform or to divert their energies into a parallel model with differences in a minor area, a sub-model could be used to redefine that area in the dissident group's terms. The sub-model would therefore offer an alternative interpretation.

Group 11: Languages as sub-models

It may be convenient, for some purposes, to consider the relationships between theoretical formulations used in a particular language as a sub-model. The differences between the concepts encountered in Indo-European languages are relatively minor, so that term equivalents pose no great problems, but should it be necessary to enrich the system by incorporating theoretical formulations from other language groups, problems could arise.

DATA TO BE INCLUDED ON EACH ENTITY

A. Concept Filing Phase [Identification or Registration] 1.

Entity Sequence Number (1)

Each new conceptual entity, of whatever type (see earlier section), receives a unique number which is the next available in a sequential list. The number therefore contains no significant digits or codes and has no meaning for classification purposes. [It may be an advantage to use the check digit technique.]

For practical purposes, it may be convenient to pre-allocate blocks of numbers to different filing centres whenever required. This avoids problems of duplication and speeds up administration. Where duplication does occur, this is eliminated at the modelling stage.

One advantage of this sequence number as a concept identifier is that it is not necessary to file a definition or conventional term at the same time. This is convenient if a new theoretical formulation has been tentatively conceived with known relationships to other concepts but with no clear definition or label yet. It avoids the need to coin doubtful neologisms in order to register the concept. In some cases it may even be an advantage to leave the term defined by its context of relationships, and not to bother attempting to find a suitable term. In which case the sequence number would be used as the only identifier until a suitable terminology for concepts in that domain can be elaborated more systematically.

2. Model Description

2.1 Model Number

The act of filing an entity is distinct from the later modelling activity. The "model number" in this case is "0". This artifice permits the definitions and the conventional terms or labels in different languages to be handled within the computer record framework together with the modelling and term allocation activity.

2.2 Sub-model Number (see model. Group '10)

Again, since entity filing is distinct from the later modelling activity, this zone is "free". It is therefore used to distinguish between

entity definitions (for which it is "0")
entity conventional labels or terms (for which it is "1")

2.3 Language (see model Group 11)

Since the definitions or the label may be given in several languages a language code is used, (e.g. English "1", French "2", etc.).

(1) Paragraph numbers refer to columns in figure 3 of the computer record layout. No attempt has been made at this preliminary stage to indicate how many character positions would be required for each zone in the record.

Descrip. A	Cross-	(v ,ur.	Model	Descrip. B	Relat.	f. iat.	Date	Date	Reten.	Status
Lang. Alt,	ref. No.		Model	Sub- Lang,	type A	type B	first	last	period	codes
1 1 2.1	3	4	5.1	5.2 5.3	5.1	6 . 2 7 . 1	7.1	7.2	7.3	8-
1 0	0	D	0	0 0						different
2 0	0	0	0	0 0						language
3 0	0	D.	0	D 0						labels
1 0	849	0	0	0 0						
* 1 0	279	D	0	0 0						
1 0	988	D	0	0 0						
2 0	434	0	0	0 0						
2 0	9321	:	0	0 0						
1 1	0	0	0	0 0						
1 2	0	0	0	0 0						
1 3	0	0	0	0 0						
2 1	0	0	0	0 0						
2 2	0	0	0	0 0						
0 0	8914	.	0	1 0						
0 0	3256	0	0	1 0						
0 0	232	0	0	1 0						
0 0	6754	0	0	25 0						
0 0	121	0	0	25 0						
0 0	4132	0	0	25 0						
1 0	0	0	0	0 0						
2 0	0	0	0	0 0						different
3 0	0	0	0	0 0						language
1 1	849	0	0	0 0						terms
1 2	279	.	0	0 0						

Outline of computer record layout (Column heading numbers refer to text section heading)

2.4 Alternatives

There are bound to be cases, for a given language, in which alternatively worded definitions (with the same meaning) are put forward. Similarly, where several conventional terms or labels referring to the same entity exist, these may also have to be filed. A simple sequential code ("1", "2", etc.) is therefore used to distinguish between successive alternatives.

3. Cross-reference

Cross-references are used during the modelling phase so that this zone is "free". It is, however, used in this phase to identify the sequence number of

other entities which use the same conventional labels as this entity (i.e. where the same label is used with a different meaning)

other entities which are defined using the same verbal definition . (but for which the definition has a different meaning]. -This may be a low-frequency or trivial case.

4. Source Code

There are several possible ways of handling information about the source of information on the entity.

- 4.1 Ignore. In a simplified system it is not necessary to include it since such information can be found in a backup card file.
- 4.2 Abbreviate. Some general code, indicating the country, the publication, or the filing group can be used.
- 4.3 Name. The name of the person, or filing organization, may be given in some abridged form (e.g. "DEUTKW" for Karl W. Deutsch).
- 4.4 Name and Support. In a more elaborate system, in which members of a discipline are expected to indicate any strong "support" or "opposition" to any new theoretical formulation, a "voting" technique may be envisaged (see page). This option could be confined to the "elders" of the profession -- or left open to all members of a profession. As "professional" activity, this might be restricted to the modelling phase.

A given member of the profession, if sufficiently aroused, could then file his support or opposition in the form "DEUTKW +" or "DEUTKW -".
- 4.5 Name and Reference. It might be thought more valuable to give not only the name but the reference to the document in which the theoretical formulation is discussed and justified.

On the question of abbreviations to document reference, one is immediately in the jungle of dispute amongst librarians, documentalists, etc. Several possibilities exist.

- 4.5.1 Use an extended bibliographical "standard" reference.
This uses a lot of space and is mainly pleasing to librarians.
- 4.5.2 Use an abbreviated reference as in "Science Citation Index" (e.g. the first four letters of the first two significant words of the title, plus the year date, issue or volume number within which pagination is consecutive, and the first significant page number - "DEUT RW -- NERV GOVE -- 1963 -- 0 - 192"].
- 4.5.3 Use a sequence number code. To avoid getting bogged down in documentation problems, a simple sequence number could be used for each publication:

either: i) referred to by the system (e.g. a complete sequence across all authors)

or: ii) referred to by the system for a given author (e.g. starting from zero for each new author).

A parallel "documentation" system would be required to decode the codes used in the approach but it might prove much tidier and practical in the long run (e.g. "DEUTKW 509") (1). The precise page numbers might be an additional requirement (e.g. "DEUTKW 509-192"). Again, as a "professional" activity, this might form part of the modelling phase.

5. Model Descriptor

This is not used during this phase.

6. Relationship Descriptor

This is not used during this phase.

*

7. Date Codes

7.1 Date first used

The date on which a theoretical formulation was first used is inserted here. If this is not supplied, the computer can automatically insert the date on which the entity was filed.

7.2 Date last used

This date is supplied as a result of general consensus by all modelling groups and is therefore not dealt with during this phase.

7 . 3 Retention period

It may be an advantage in this phase to tag some entities of unknown value so that they will automatically be dropped from the system after a certain period unless some contrary instruction is received in the meantime. Different retention periods can be used according to the status of the source.

(1)For a very useful discussion of this approach to documentation, see Jacques E.J. Halkin, "Proposal and wishes for an open structure in the communication of information." Scheduled for publication in: A.I. Mikhailov (Ed.) The Theoretical Problems of Information Retrieval Systems. (The Hague, International Federation for Documentation, 1971)

8. Status

For administrative purposes it is convenient to have a zone in which codes may be used to indicate that the entity is "under consideration", "of doubtful value", "no longer used", etc.

9. Text

The words or text used for:

the conventional terms or labels
the definitions

would be inserted into this zone. This zone could also be used for any special comments which might be usefully added.

B. ' Concept Coding Phase (Modelling or Classification)

Many of the zones discussed above are used in this phase but for a different purpose or in order to establish computer records distinct from those created during the earlier phase or by other modelling groups.

1. Entity Sequence Number

This is repeated for each new relationship established within a model and is of course the same as that used in filing the identity in A.1.

2. Model Descriptor

2.1 Model number

As discussed elsewhere (see page), each modelling group receives a unique number (e.g. "362") which identifies the system of relationships which are elaborated and filed, while at the same time distinguishing it (at computer level) from any other systems.

There is some argument for attaching special significance to particular digits of the model number with a view to clarifying a hierarchy of models or, at least, showing a relationship between models. In other words, at this level a U.D.C.-type approach might be used so that "political science" models are all identified by "32N" and "anthropology" models by "39N". This is probably a temptation to be resisted however, since it has some theoretical implications which are better contained within models. In which case a simple sequential list should be established from which the next available model number could be taken.

2.2 Sub-model number

This is a zone to be used by a modelling group whenever a level of dissent is encountered so that alternative sub-models within the general model can be satisfactorily handled and identified. Normally, in the absence of sub-models, this would be "0".

2.3 Language

Since the relationship between concepts is supposedly language independent, this zone should normally be "0".

There are, however, cases where relationships are identifiable in one language but absent, ridiculous, or ambiguous in a second. In such cases it may be convenient to use this zone for a form of language-dependent sub-model.

2.4 Alternatives

This zone is not used in this phase and must be "0" (to permit identification of the term records in the next phase where it is non-zero).

3. Cross-references

This zone supplies the main means by which the relationship of this entity to other entities is indicated for the particular model indicated in 2.1. The sequence number of the other entity is indicated here. In effect, every such "relationship" gives rise to a new computer record (see figure 3).

The type of relationship is either implicit, because of the model used, or is described in 6 and 7.

4. Source Code

Depending on the method chosen (see A.4.1, A.4.2, A.4.3, A.4.4, and A.4.5), the source coding would probably either be allocated during the concept filing phase with nothing in this phase, or in this phase with nothing in the previous phase. In the most sophisticated system, it might however be desirable to give:

source coding for the entity in the concept filing phase
source coding for individual relationships within a model,
during the modelling phase.

Source coding during the modelling phase might be particularly helpful in the administrative work of elaborating a model, since it permits members of a modelling group, working independently and in isolation, to "vote" on the insertion or deletion of particular relationships (see A.4.5). Such a postal vote system would be particularly helpful in clarifying with precision just what was under discussion at any point in time.

5. Model Descriptor

This zone is used to indicate which model is to be considered at the entity cross-references in 3.

In a simplified system this zone would not be required because the assumption would be made that each model was totally isolated from other models*

In a more sophisticated system however, there is need for a means of expressing relationships between parts of models. For example, it may be that in a certain domain two models are identical or that one forms a subset of the other. In such a case there is little need to duplicate all the relationships in the second model, provided cross-reference between the models is possible.

5.1 Model number

As for 2.1, but the model is only to be entered at the entity in which the cross-reference in 3 refers.

5.2 Sub-model number

As for 2.2, but again is only to be entered at the entity to which the cross-reference in 3 refers.

5.3 Language

As for 2.3, but again is only to be entered at the entity to which the cross-reference in 3 refers.

5.4 Alternatives

Not used. (This zone may even be omitted entirely.)

6. Relationship Descriptor

This zone is used to describe the relationship constituted by the link between this entity and that cross-referenced in 3. Two basic types of relationship descriptors may be distinguished.

6.1 Relationship descriptor A

This is used to give an indication of the relative levels of the two entities related (e.g. class and member], directions for flow (e.g. from or to), etc. These are used, for example, to indicate any hierarchical relationships. These codes and the cross-reference in 3 are all that is required for a graph-theoretical analysis of the network of concepts.

It is here that any "see other" code would be inserted.

It is also important to indicate the type of relationship between two entities (see page):

```
-- ' logical"(i.e. B includes A, etc.)
      consistency (con'tradiction/support)
      time(precedes/follows)
      cybernetic (information exchange)
      responsibility (flow of decisions)
-   etc. '
```

This is an indication of what is flowing or the nature of the relationship. It does not seem feasible to predetermine the possible types of relationship which might be required. The technique which can be adopted is therefore to use a simple numeric code --the next available in a sequential list -- for each new type of relationship with which a modelling group wishes to work.

The arrangement of this zone could be left up to the modelling group. It is desirable that standard codes should be developed to facilitate graph-theoretical analyses and that a standard code system should be used to denote types of relationships (e.g. "321" where the numbers have no special significance].

6,2 Relationship descriptor B

This is used for evaluation descriptors. In other words the codes used here supply some form of ranking to the relationship described in 6.1 (e.g. some measure of relative importance (within the model), some measure of degree of relativity, etc.)

It is in this zone that the degree of consensus on the characterization of the concept by the discipline could be coded.

The zone may even be used to carry quantitative information on the size of flow represented by the relationship and also its periodicity, if relevant.

the arrangement of this zone could be left up to the modelling

It is however desirable that a standard form should be developed -- even if exceptions to it are frequent.

7. Data codes

7.1 Date first used

This may be used to indicate the date each relationship between entities was first noted, or alternatively the computer can automatically insert the date on which the relationship was first filed.

7.2 Date last used

This date may be used when the relationship is finally rejected as invalid or unacceptable.

7.3 Retention period

This zone may be used by members of a modelling group to communicate with one another. A member may submit "trial balloon" relationships, with a very short (one-cycles) retention period so that others can "see how it looks". Once agreed, the retention period can be set so that relationships periodically come up for review.

8. Status code

For modelling group administration purposes, it is convenient to have a zone which may be used to indicate that the relationship is "under consideration", "a tentative proposal", "a firm proposal", "agreed by the group", "required priority attention", etc.

9. Text

Normally a relationship record should require no text. There is however no reason why this zone should not be used for any text comments on a relationship which may seem significant to the modelling group.

C. Term Allocation Phase

1. Entity sequence number

Required as before.

2. Model descriptor A

2.1 Model number

Required as before. A term can only be authoritatively allocated within the modelling group. It is Utopian to expect that consensus can be consistently achieved between modelling groups on a unique authoritative term for the entity to which they all refer in their different ways.

2.2 Sub-model number

This should normally be zero, since it will probably be easier to achieve consensus on a term between model and sub-model than between model and model.

2.3 Language

Required as before for each language version of the authoritative term.

2.4 Alternatives

This must be "1" or greater to distinguish the term records from the relationship records. If alternative authoritative terms are required in a given language the zone would be used to distinguish between them.

3. Cross-reference

Normally this would be "0". It may however be necessary to indicate other entities using the same term (but obviously "with a different meaning).

4. 'Source code

There may be some cases where it is important to indicate the document in which the justification for the unique authoritative term is urged.

5. Model descriptor B

May be required if the cross-reference to a use of the same term in a different model is needed. -

6. Relationship descriptor

Not required.

7. Data codes

7.1 Date first used

This may be used to indicate the date the term was first used, or alternatively the computer can automatically insert the date on which the term was first filed.

7.2 Date last used

Terms fall from favour. The last date of use can be indicated here.

7.3 Retention period

May be used as in B.7.3.

8. Status code

May be used as in B.8

9. Text

The words used in the authoritative term are inserted into this zone. Alternatively, the equivalent decimal coding could be inserted, if desired.

LIMITATION OF SCOPE AND SOURCES OF CONCEPTS

1. Scope

The design of the system is sufficiently general that it could be used to order theoretical formulations in any area of Knowledge. Such broad coverage would clearly be impracticable, and probably even undesirable, in the foreseeable future.

It is useful to re-emphasize that the proposal is not concerned with the areas covered by social science documentation as there are many such documentation projects. The UNISIST report mentions the parallel programs proposed by such bodies as the International Council of Social Sciences and the International Committee for Social Sciences Documentation. There are numerous equivalent projects at the national level. The object here is to concentrate on theoretical formulations which may or may not be mentioned in a given collection of documents.

The priorities proposed would be based on three dimensions:

- i) commencing with the more abstract formulations and then moving to the more specific or concrete
- ii) commencing with formulations of interest to several social science disciplines and then moving to those common to several schools of thought, and finally to those current within one school of thought only. (The suggestion is that an effort should be made to elaborate the significance of "inter-", "multi-" or "trans-disciplinary" concepts as a priority area of study with respect to knowledge analogous to the focus on international relations as opposed to national level activities. The degree of interdisciplinarity of a concept is a valuable means of determining priorities(1).

(1) See: OECD, Centre for Educational Research and Innovation. Inter-

I

disciplinarity; problems of teaching and research in the universities. Paris, OECD, 1972, 321 p.

- iii) commencing with theoretical formulations before going on eventually to methods and supporting data

This does not of course preclude any modelling group from concentrating solely on the formulations of its own school of thought. The main concern however should be to ensure that the system reflects the general framework of theoretical formulations. Highly specialized formulations should not clutter up the modelling activity. Little effort should be made to include minutiae about particular social entities which have not been . reflected in more general formulations -- unless such minutiae represent unique evidence of the need for new formulations. The system should be compact and easy to use rather than large and unwieldy as are most documentation systems.

2. Sources

Guidance in limiting scope can be obtained by concentrating in the light of the above priorities on concepts mentioned in such publications as

- 2.1 David L. Sills (Ed.), International Encyclopedia of the Social Sciences, Macmillan, 1968. .
- 2.2 Julius Gould and W.L. Kolb (Ed.), A Dictionary of the Social Sciences (compiled under the auspices of Unesco) New York, Free Press, 1964.
- 2.3 UNESCO, Main Trends of Research in the Social and Human Sciences. Paris, Unesco, (Part one: social sciences, 1970, 819 p.; Part two: human sciences, 1972). Also in French edition.
- 2.4 International Committee for Social Sciences Documentation. International bibliography of the social sciences. London, Tavistock, 4 annual volumes (sociology, political science, economics, social and cultural anthropology).
- 2.5 Key textbooks in each discipline.
- 2.6 Specialized multi-lingual dictionaries and glossaries, such as:
 - Günter Haenich. Dictionary of International relations and politics; systematic and alphabetical in four languages (German/English/French/Spanish). Elsevier, 1965.
 - This dictionary has 5778 terms with equivalents in the four languages.
 - I. Paenson. English/French/Spanish/Russian Systematic Glossary of Select Economic and Social Terms. Oxford, Pergamon, 1964.
 - Attempts to present a system of inter-related concepts which reflect a vertical hierarchy and are presented within a continuous text in a systematic exposition of a given subject.
- 2.7 Institute for Scientific Information. Social Sciences Citation Index. Philadelphia, (included in the SSCI are three separate but related indexes of different periodicity covering the literature of the specified calendar year. Price in 1973: \$1,250(sic)).

2.8 International social science organizations.

A preliminary count indicates that possibly some 30 such bodies could contribute in some way to the project.

CONCEPT NOTATION IN DOCUMENTS

It has been stressed that this project does not require a complex notation system since each concept is represented by a single sequence number, plus an indication of the model number in question, if required. Nevertheless, since one object of this approach is to permit scholars to refer with precision to a particular concept in their papers, a standard method of indicating such a concept in print is required.

A similar problem arises in the natural sciences in distinguishing between different isotopes of the same atom (i.e. cases where slightly different versions of the same atom exist due to differences in atomic weight), where the same symbol does not distinguish between isotopes. The solution adopted is to indicate the atomic weight as a superscript to the standard symbol.

In the case of concepts, represented in print by the same word, one solution would be to use the sequence number of the concept as superscript to the word:

e.g. démocracy⁺²⁵¹ 'démocracy⁺⁹⁴²

To avoid confusion with bibliographical references, the number could perhaps be preceded by an asterisk.

There is a strong temptation to adopt a technique for uniquely identifying concepts similar to that of the International Standard Book Numbering [ISBN] system now used on the reverse of all recent book title pages] to give a unique code to each book. This number consists of 10 digits made up of the following parts:

- group identifiers (i.e., national, geographical, language or other convenient group]. An "agency" coordinates the allocation of numbers within each group e.g., one for Anglo-American publications ("0"), one for UN system publications, etc. The group identifier is allocated by an international standard book numbering agency (in formation]. (This could be considered as a concept filing centre identifier allocated by some loose coordinating body.)
- book publisher identifiers. The publisher identifier is allocated internally within the group by the group agency. (This could be considered as an accredited concept filing source identifier allocated with respect to the filing centre for which it locates new conceptual entities]
- book title identifiers. A block of sequence numbers is reserved for each publisher to permit him to select the next available for the next book. (This could be considered as a block of sequence numbers for concepts, so that each accredited source can select the next number as each new concept is identified.)
- ~ check digit. This ensures that the code has been correctly transcribed and input to the computer. A computer pre-generated list of "available" sequence numbers incorporates this digit (which is calculated on a modular 11 with weights 10-2, using X in lieu of 10 where 10 would occur as a check digit).

The total length is 10 digits, but the three identifiers only total 9 digits. In order to avoid wastage of numbers or lack of sufficient numbers, publishers

with a large book output (of which there are few] have a two or three digit identifier so that the title identifiers can use six or five digits. A small publisher (of which there are many] has a five or six digit identifier so that the title identifier can use two or three digits. The publisher identifier is therefore selected on the basis of his output using from two to six digits as required. Hyphen separators are used.

The temptation to use this system should however be resisted. While the significance attached to the digits is only "administrative" and has no "theoretical" implications, problems of overflowing the allocated blocks are bound to occur. The system will "bulge" in unpredictable areas as the U.D.C. has done. It is also questionable whether so much significance should be placed on the source which, once the concept has been incorporated, will quickly become irrelevant within the network of other related concepts from other sources.

REPRESENTATION OF CONCEPT NETWORKS USING GRAPH THEORY

This project is concerned with the collection of entities and the indication of relationships, if any, between those entities. Expressed in these general terms, the techniques of graph theory may be used in this project. Graph theory is concerned with the "arcs" [links or relationships] between "nodes" (entities] and the various structural properties of the network so constituted.

It can be of great assistance in dealing with, a broad range of combinatorial problems which occur in various economic, sociological or technological fields. It is, perhaps, that aspect of the theory of sets which can produce the most fruitful results not only for the pure mathematician, the engineer, and the organizer, but also for the biologist, the psychologist, the sociologist and many others. Graphs can be used to represent structures such as: a network of roads, an electrical circuit, communication in a group, a complex chemical molecule, circulation of documents in an organization, kinship structures, .etc. (1)

Its use in connection with relations between more abstract social entities such as organizations and nations is much less frequent(23. Its use for handling

-
- (1] A. Kaufman. Graphs, dynamic programming and finite games. N.Y., Academic, 1967.
Claude Berge, Théorie des graphes et ses applications. Paris, Dunod, 1958, 277 p. •
- (2] Claude Flament. Théorie des graphes et structures sociales. Paris, Houton, 1965 (English edition, Prentice-Hall].
J.Clyde Mitchell (Ed]. Social Networks in Urban Situations, Manchester U.P., 1969
Norman Schofield. A topological model of international relations. (Paper presented to Piece Research International meeting, London, 1971]
George M. Seal et al. System linkages among women's organizations. Department of Sociology and Anthropology, Iowa State University, 1967.
- Robert O. Anderson. A sociometric approach to the analysis of inter-organizational relationships. Institute for Community Development and Services, Michigan State University, 1969.
- D. Cartwright. The potential contributions of graph theory to organization theory. In: M. Haire (Ed.] Modern Organization Theory, Wiley, 1959.

psycho-social abstractions appears to be even rarer (1).

The image of a 'network or web of ideas' to represent a complex set of inter-relationships in a sphere of knowledge, and particularly culture, is a fairly familiar one. This use of 'network', however, is purely metaphorical and is very different from the notion of a network of concepts as a specific set of linkages among a defined set of concepts, with the additional property that the characteristics of these linkages as a whole may be used to interpret the semantic significance of the concepts involved.

Some features of concept networks

Points 1 to 3 below are concerned with the shape of the network, 4 to 8 with interactions within the network.

1. Centrality. A measure (in topological not quantitative terms) of the extent to which a given theoretical entity (e.g. a concept) is directly or indirectly "related" via links to other entities i.e., the extent to which it is "distant" from another entity. One can speak of a "key" concept or of a concept being "central" to the concerns of a particular discipline. It may also be considered a measure of the degree of "isolation" of the entity. A systematic analysis of the centrality of theoretical entities could indicate where new concepts are necessary to bridge conceptual gaps and link isolated domains.
2. Coherence. A measure of the degree of "interconnectedness" or "density" of a group of concepts. This may be considered as the degree to which a system of concepts is "complete". Differences in density would reflect the tendency for more highly coherent concept systems to appear more self-reinforcing in comparison to less organized parts of the network. In some respects this is an indication of the degree of "development" of a group of concepts.
3. Range. Some concepts are directly related to many other concepts, others to very few. The range of a concept is a measure of the number of other entities to which it is directly related.

Range could be considered an indication of the "vulnerability" of a concept, to the extent that a high range concept would be less vulnerable to attack than a low range concept, since it has more bonds anchoring it to its

(1) In the field of documentation a thesaurus may be represented "graphically" but more for the visual presentation facility than for any graph theoretic possibilities. For example: the "genetic maps" of the U.S. Armed Services Technical Information Agency (ASTIA), the concentric circle diagram of the Technische Dokumentatie - en Informatie Centrum voor de Krijgsmacht (TDCK, The Hague), the arrow diagrams used by EURATOM and the Bureau d'etudes van Dijk in Brussels (see Figure 4). See also the computer established "association maps" of Lauren B. Doyle. (Indexing and abstracting by association. American Documentation, October, 1962). See Also: Kurt Lewin. The Principles of Topological Psychology. N.Y., McGraw-Hill, 1936; E. Zierer. The theory of graphs in linguistics. The Hague, Mouton, 1970, 62 p.; R. Quillan. Semantic memory. In: M. Minsky (Ed.). Semantic Information Processing. Cambridge, M.I.T., 1968, p. 225-270; R.B. Banerji. A language for the description of concepts. Unpublished paper, System Research Center, Case Institute of Technology, 1964.

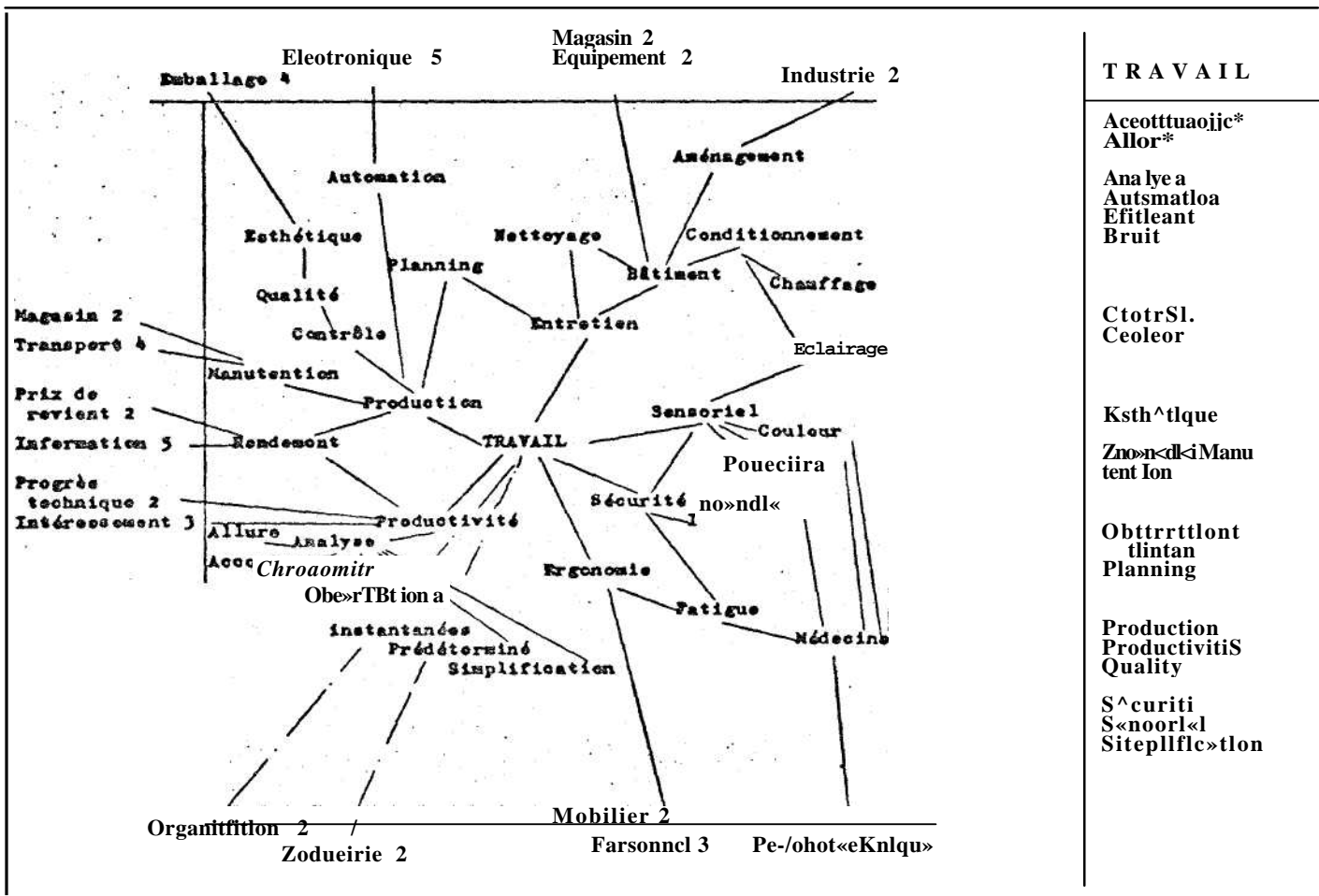


Figure 4. Arrow diagram used in special indexes

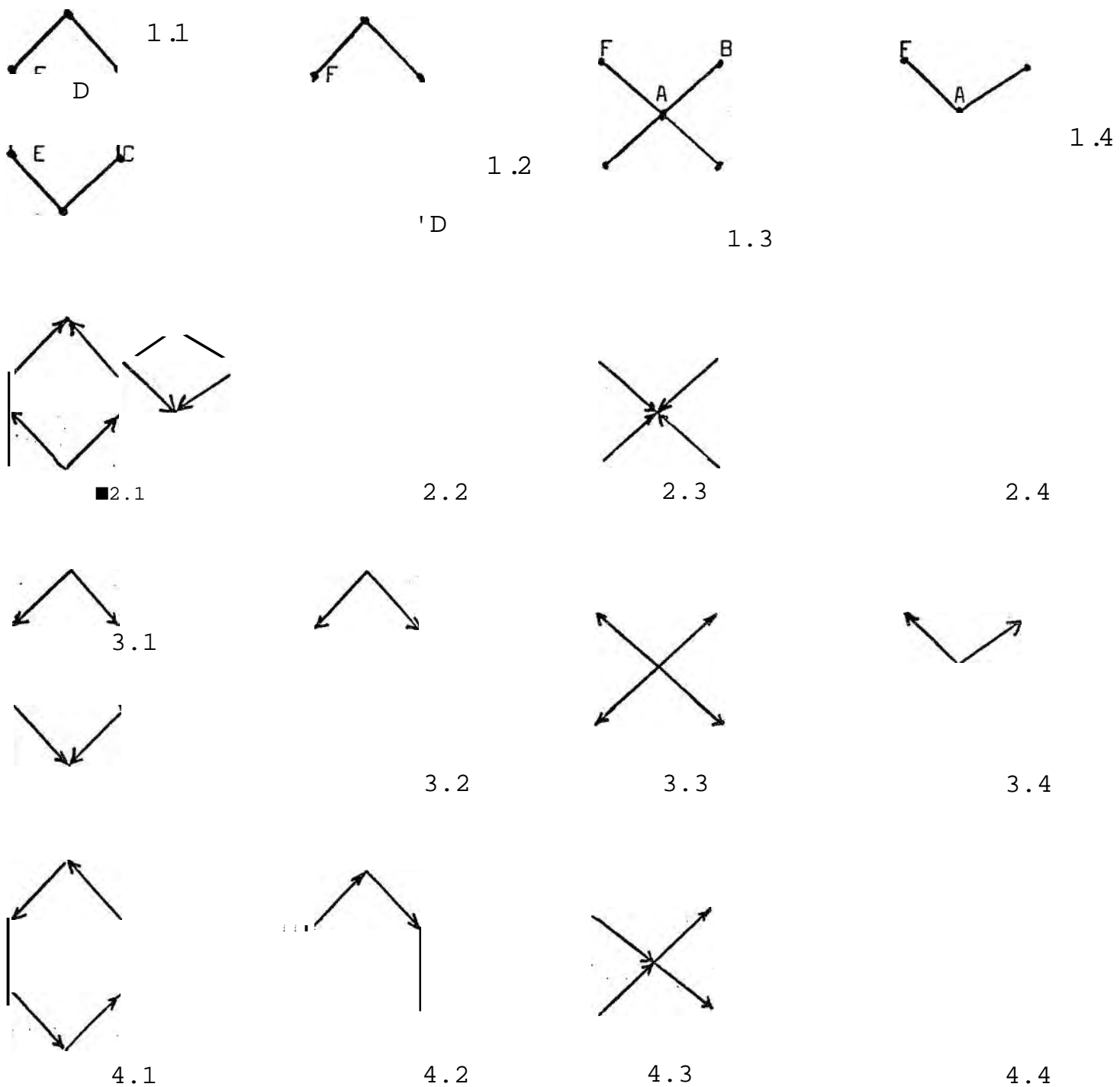


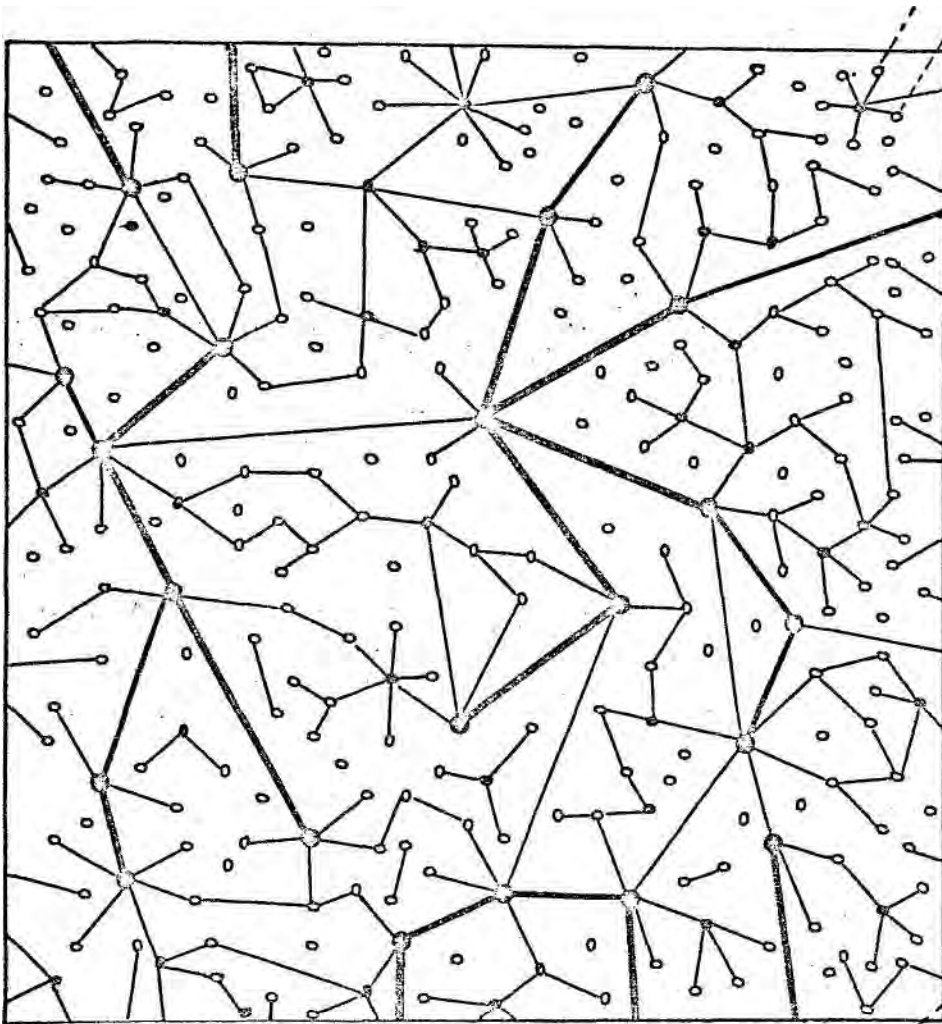
Figure 5 Examples of simple networks of entities

- a] In the non-directed examples of group CD, A is the central concept in (1.3), A and D in [1.4], A and F in (1.2). In (1.1), there is no central concept.
- b] In group (1), peripheral concepts are D and C in (1.2),- B, C, E and F in (1.3); B, C and F in (1.4). There are no peripheral concepts in (1.1).
- c] In group (1), the range of A in (1.3) is 4, in (1.4) it is 3.
- d] In group (1), the reachability of A in (1.1) and (1.2) is 3, in (1.3) it is 1, and in (1.4) it is 2.
- e] In all the directed examples of group (2), A is the central concept with at least B and F as direct component concepts. In all except (2.3), there are even sub-sub-components of A.
- f] In all the directed examples of group (3), A is the central concept but only as a common sub-component. D is also a common sub-component in (2.1).
- g] In all the directed examples of group (4), there is a chain of component/sub-component links. In (4.1), this is continuously forming a loop. In (4.2) and (4.4), C is the major concept. In (4.3), A is the central concept but only by having F and E as sub-components and being itself a common sub-component to B and C.

The above features are all evident, almost to the point of being trivial. But most cases of interest are likely to be much more complex, with many nested levels of concepts and cross-linking relationships. These may be examined by matrix analysis techniques, particularly using computers (to which the proposed record layout is suited)(1). Computer programs exist to detect properties of networks.

A more complex example is illustrated by Figure 6. There is shown the manner in which two different models or conceptual structures might interlink the same concepts to form two very different patterns - which may be analysed.

-
- [1] C. Berge. The Theory of Graphs and its Applications. London, Methuen, 1962.
 - C. Flament. Applications of Graph Theory to Group Structure. Englewood-Cliffs, Prentice-Hall, 1963
 - F. Harary and R.Z. Norman. Graph Theory as a Mathematical Model in Social Sciences. Ann Arbor, University of Michigan, 1953.
 - F. Harary, R.Z. Norman and D. Cartwright. Structural Models: an introduction to the theory of directed graphs. N.Y., Wiley, 1965.



Network of concepts
'under Model A
(same concepts as for
Model B, but different
pattern of links)

	Concepts	Links
Primary	22	14
Secondary	33	8
Tertiary	147	228
isolates	58	n.a.
Isolated groupe	4	n.a.

i.e. network is more
integrated, with fewer
isolates; but some
primary coord" inating
entities are only related
via tertiary links.

S'Q

■ Network of concepts
under Model B
(same concepts as for
Model A, but different
pattern of links)

	Concepts	Links
Primary	8	9
Secondary	12	0
Tertiary	140	130
Isolates	100	n.a.
Isolated groups	34	n.a.

i.e. network is less
integrated, with more
isolates; but all primary
coordinating entities are
related via primary links

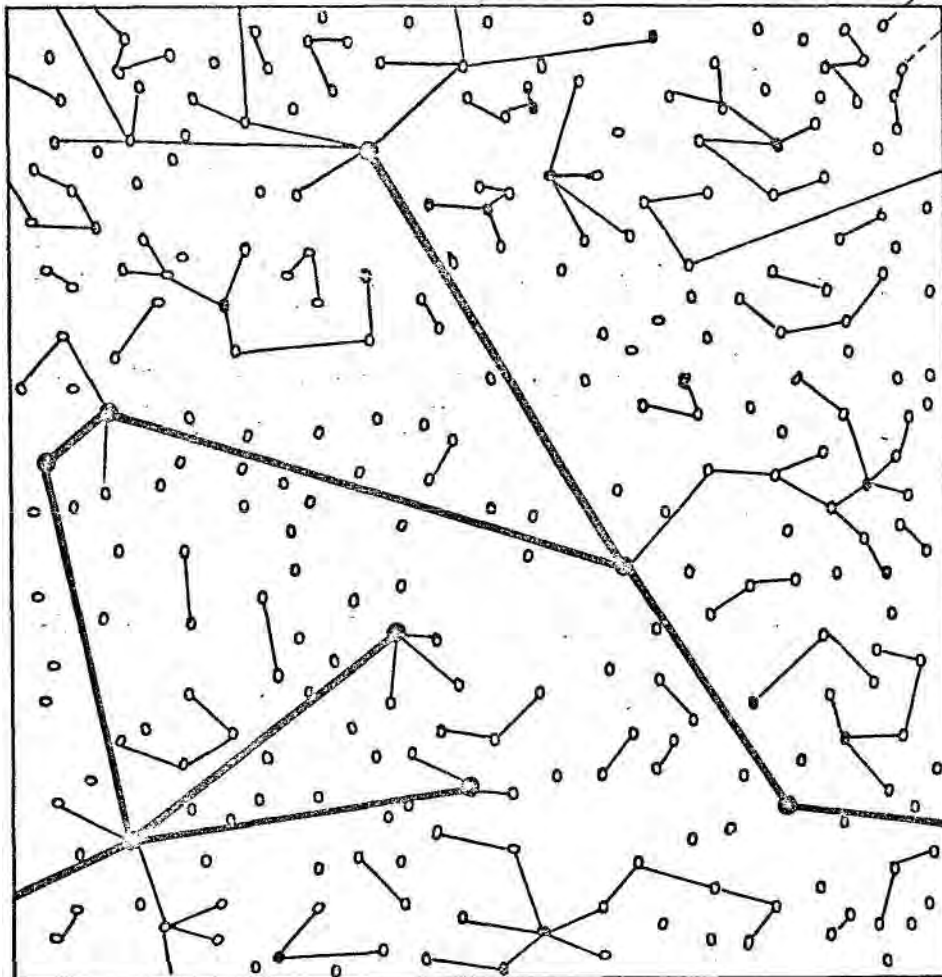


Figure 6.
Illustration of how
one part of the
network of concepts
might appear with
respect to two diff-
erent models or
schools of thought
(or, alternatively
at two different
dates).

USE OF INTERACTIVE GRAPHIC DISPLAY TECHNIQUES

Description

The suggestion has been made (see previous section) that structuring the relationship between theoretical entities [concepts, propositions, problems, etc) could best be accomplished using graph theory methods. There are three disadvantages to this approach:

- graphic relationships are tiresome and time-consuming to draw (and are costly if budgeted as "art work").

once drawn, there is a strong resistance to updating them [because of the previous point) and therefore they quickly become useless.

when the graph is complex, multidimensional, and carries much information, it is difficult to draw satisfactorily in two dimensions. The mass of information cannot be filtered to highlight particular features - unless yet another diagram is prepared.

These three difficulties can be overcome by making use of what is known as "interactive graphics" [1). This is basically a TV screen attached to a computer. The user sits at a keyboard in front of the screen and has at his disposal what is known as a light-pen [or some equivalent device) which allows him to point to elements of the network of concepts displayed on the screen and instruct the computer to manipulate them in useful ways. In other words the user can interact with the representation of the conceptual network using the full power of the computer to take care of the drudgery of

drawing in neat lines

making amendments

displaying only part of the network so that the user is not overloaded with "relevant" information

In effect the graphics device provides the user with a window or viewport onto the network of concepts. He can instruct the computer, via the keyboard, to:

1. move the window to give him, effectively, a view onto a different part of the network - another conceptual domain
2. introduce a magnification so that he can examine [or "zoom in" on) some detailed sections of the network
3. introduce diminution so that he can gain an overall view of the structure of the conceptual domain in which he is interested
4. introduce filters so that only certain types of relationships and entities are displayed - either he can switch between models or he can impose

(1) This term is used widely to cover both the more common "alphascopes", which can display letters and numbers on predetermined lines, and the "vector displays" with light-pen facility, which can also generate lines and curves. It is the latter device which is discussed here. See, for example:

See: Ivan Sutherland. Computer displays. Scientific American, 222, June 1970, p. 56-8.

Interactive graphics in data processing. IBM Systems Journal, 7, 3 and 4, 1968, whole double issue.

Computer Graphics 1970; and international symposium. Brunei University, 1970, 3 vols.

restrictions on the relationships displayed within a model, i.e. he has a hierarchy of filters at his disposal.

5. modify parts of the network displayed to him by inserting or deleting entities and relationships. Security codes can be arranged to that (a) he can modify the display for his own immediate use without permanently affecting the basic store of data, (b) he can permanently modify features of the model for which he is a member of the responsible body, (c) and so on.
6. supply text labels to features of the network which are unfamiliar to him. If necessary he can split his viewport into two Cor more] parts and have the parts of the network displayed in one (or more) part(s). He can then use the light pen to point to each entity or relationship on which he wants a longer text description (e.g. the justifying argument for an entity or the mathematical function, if applicable, governing a relationship, and have it displayed in an adjoining viewport.)
7. track along the relationships between one entity and the next by moving the viewport to focus on each new entity. In this way the user moves through a representation of "semantic space" with each move, changing the constellation of entities displayed and bringing new entities and relationships into view.
8. move up or down levels or "ladders of abstraction". The user can demand that the computer track the display (see point 7) between levels of abstraction, moving from sub-system to system, at each move bringing into view the semantic context of the system displayed.
9. distinguish between entities and relationships on the basis of user-selected characteristics. The user can have the "relevant" (to him) entities displayed with more prominent symbols, and the relevant relationships with heavier lines.
10. select an alternative form of presentation. Some users may prefer block diagram flow charts, others may prefer a matrix display, others may prefer Venn diagrams Cor "Venn spheres" in 3 dimensions) to illustrate the relationship between entities. These are all interconvertible (e.g. the Venn circles are computed taking each network node as a centre and giving a radius to include all the sub-branches of the network from that node).
11. copy a particular display currently on the screen. A user may want to keep a personal record of parts of the network which are of interest to' him. (He can either arrange for a dump onto a tape which can drive a graph plotter, a microfilm plotter, or copy onto a videocassette, or, in the future, obtain a direct - photocopy.)
12. arrange for a simultaneous search through a coded microfilm to provide appropriate slide images or lengthy text (which can in its turn be photo copied).
13. simulate a three-dimensional presentation of the network by introducing an extra coordinate axis.
14. rotate a three-dimensional structure (about the X or Y axis) in order to heighten the 3-D effect and obtain a better overall view "around" the structure.
15. simulate a four-dimensional presentation of the network by using various techniques for distinguishing entities and relationships (e.g. "flashing" relationships at frequencies corresponding to their importance in terms of the fourth dimension.)

16. change the speed at which the magnification from the viewport is modified as a particular structure is rotated.
17. simulate the consequences of various changes introduced by the user in terms of his conditions. This is particularly useful for cybernetic displays.
18. perform various topological analyses on particular parts of the network and display the results in a secondary viewport (e.g., the user might point a light-pen at an entity and request its centrality or request an indication of the interconnectedness of a particular domain delimited with the light pen.) .

In every current use of interactive graphics there is some notion of geometry and space, but the geometry is always the three-dimensional conventional space. There is no reason why "non-physical spaces" should not be displayed instead -and this is the domain of topology. The argument has been developed by Dean Brown and Joan Lewis (1)

"Both geometry and topology deal with the notion of space, but geometry's preoccupation with shapes and measure is replaced in topology by more abstract, less restrictive ideas of the qualities of things... Being more abstract and less insistent on fine points such as size, topology gives a richer formalism to adapt as a tool for the contemplation of ideas....

Concepts can be viewed as manifolds in the multidimensional variate space spanned by the parameters describing the situation. If a correspondence is established that represents our incomplete knowledge by altitude functions, we can seek the terrae incognitae, plateaus, enclaves of knowledge, cusps, peaks, and saddles by a conceptual photogrammetry. Exploring the face of a new concept would be comparable to exploring the topography of the back of the moon. Commonly heard remarks such as "Now I'm beginning to get the picture" are perhaps an indication that these processes already play an unsuspected role in conceptualization....

By sketching tentative three-dimensional perspectives on the screen and "rotating them on the tips of his fingers", one internalizes ideas non-verbally and acquires a sensation of sailing through structures of concepts much as a cosmonaut sailing through constellations of stars.

Such new ways of creating representations break ingrained thought patterns and force re-examination of preconceived notions. A mapping is a correspondence is an analogy. Teaching by analogy, always a fertile device, can be carried out beautifully by topological means.... Topological techniques are useful at even the most advanced levels of scientific conceptualization...."

The fundamental importance of interactive graphics, in whatever form, is its ability to facilitate understanding. Progress in understanding is made through the development of mental models or symbolic notations that permit a simple representation of a mass of complexities not previously understood. There is nothing new in the use of models to represent psycho-social abstractions. Jay

(1) Brown, Dean and Lewis, Joan. The process of conceptualization; some fundamental principles of learning useful in teaching with or without the participation of computers. Educational Policy Research Center, Stanford Research Institute, Menlo Park, California, p. 16-18

Forrester (1), making this same point with respect to social systems, states

"Every person in his private life and in his community life uses models for decision making. The mental image of the world around one, carried in each individual's head, is a model. One does not have a family, a business, a city, a government, or a country in his head. He has only selected concepts and relationships which he uses to represent the real system. The human mind selects a few perceptions, which may be right or wrong, and uses them as a description of the world around us. On the basis of these assumptions a person estimates the system behaviour that he believes is implied.... The human mind is excellent in its ability to observe the elementary forces and actions of which a system is composed. The human mind is effective in identifying the structure into which separate scraps of information can be fitted. But when the pieces of the system have been assembled, the mind is nearly useless for anticipating the dynamic behaviour that the system implies. Here the computer is ideal. It will trace the interactions of any specified set of relationships without doubt or error. The mental model is fuzzy. It is incomplete. It is imprecisely stated. Furthermore, even within one individual, the mental model changes with time and with the flow of conversation. The human mind assembles a few relationships to fit the context of a discussion. As the subject shifts, so does the model. Even as a single topic is being discussed, each participant in a conversation is using a different mental model through which to interpret the subject. And it is not surprising that consensus leads to actions which produce unintended results. Fundamental assumptions differ but are never brought out into the open."

These structured models have to be applied to any serially ordered data in card files, computer printout or reference books to make sense of that data. Is there any reason why these invisible structural models should not be made visible to clarify differences and build a more comprehensive visible model? The greater the complexity, however, the more difficult it is to 'use mental models. For example, in discussing his examination of an electronic circuit diagram, Ivan Sutherland writes(2):

"Unfortunately, my abstract model tends to fade out when I get a circuit that is a little bit too complex. I can't remember what is happening in one place long enough to see what is going to happen somewhere else. My model evaporates. If I could somehow represent that abstract model in the computer to see a circuit in animation, my abstraction wouldn't evaporate. I could take the vague notion that "fades out at the edges" and solidify it... I could analyze bigger circuits. In all fields there are such abstractions. We haven't yet made any use of the computer's capability to "firm up" these abstractions. The scientist of today is limited by his pencil and paper and mind. He can draw abstractions, or he can think about them. If he draws them, they will be static, and if he just visualizes them they won't have very good mathematical properties and will fade out. With a computer, we could give him a great deal more. We could give him drawings that move, drawings in three or four dimensions which he can rotate, and drawings with great mathematical accuracy. We could let him work with them in a way that he has never been able to do

(1) Jay Forrester. World Dynamics. Cambridge, Mass. Wright-Allen, 1971, p.14-15.

(2) Computer graphics. Datamation, May 1966, p.22-27.

before. I think that really big gains in the substantive scientific areas are going to come when somebody invents new abstractions which can only be represented in computer graphical form."

IMPLICATIONS OF COMPUTER AUGMENTATION OF INTELLECT

There are important intellectual implications emerging from work on advanced computer systems. Of particular interest is the work of Douglas Engelbart's team at the Center for Augmentation of Human Intellect (Stanford Research Institute) which is the centre for the U.S. ARPA Data Network [which links the computers of major universities in the U.S.A.]. Engelbart has worked on the means of creating an intellectual workshop to facilitate interaction between conceptual structures(1). He considers that

"Concepts seem to be structurable, in that a new concept can be composed of an organization of established concepts and that a concept structure is something which we might try to develop on paper for ourselves or work with by conscious thought processes, or as something which we try to communicate to one another in serious discussion.... A given structure of concepts can be represented by any of an infinite number of different symbol structures, some of which would be much better than others for enabling the human perceptual and cognitive apparatus to search out and comprehend the conceptual matter of significance and/or interest to the human.

But it is not only the form of a symbol structure that is important. A problem solver is involved in a stream of conceptual activity whose course serves his mental needs of the moment. The sequence and nature of these needs are quite variable, and yet for each need he may benefit significantly from a form of symbol structuring that is uniquely efficient for that need.

Therefore, besides the forms of symbol structures that can be constructed and portrayed, we are very much concerned with the speed and flexibility with which one form can be transformed into another, and with which new material can be located and portrayed.

We are generally used to thinking of our symbol structures as a pattern of marks on a sheet of paper. When we want a different symbol-structure view, we think of shifting our point of attention on the sheet, or moving a new sheet into position.

With a computer manipulating our symbols and generating their portrayals to us on a display, we no longer need think of our looking at the symbol structure which is stored -- s we think of looking at the symbol structures stored in notebooks, memos, and books. What the computer actually stores need be none of our concern, assuming that it can portray symbol structures to us that are consistent with the form in which we think our information is structured,

A given concept structure can be represented with a symbol structure that is completely compatible with the computer's internal way of handling symbols, with all sorts of characteristics and relationships given explicit identifications that the user may never directly see. In fact, this struc-

- (1) The following extracts are from
Engelbart, D.C., Augmenting Human Intellect; a conceptual framework.
Menlo Park, Stanford Research Institute, 1962, p. 34-37 CAFOSR-3223)

turing has immensely greater potential for accurately mapping a complex concept structure than does a structure an individual would find it practical to construct or use on paper.

The computer can transform back and forth between the two-dimensional portrayal on the screen, of some limited view of the total structure, and the aspect of the n-dimensional internal image that represents this "view". If the human adds to or modifies such a "view", the computer integrates the change into the internal-image symbol structure (in terms of the computer's favored symbols and structuring) and thereby automatically detects a certain proportion of his possible conceptual inconsistencies.

Thus, inside this instrument (the computer) there is an internal-image, computer-symbol structure whose convolutions and multi-dimensionality we can learn to shape to represent to hitherto unattainable accuracy the concept structure we might be building or working with. This internal structure may have a form that is nearly incomprehensible to the direct inspection of a human (except in minute chunks)".

These insights have been incorporated into the design of an operational computer system which is now being developed so that it will be possible to use computer •devices as a sort of

"electronic vehicle with which one could drive around with extraordinary freedom through the information domain. Imagine driving a car through a landscape which, instead of buildings, roads, and trees, had groves of facts, structures of ideas, and so on, relevant to your professional interests? But this information landscape is a remarkably organized one; not only can you drive around a grove of certain arranged facts, and look at it from many aspects, you have the capability of totally reorganizing that grove almost instantaneously. You could put a road right through the center of it, under it, or over it, giving you, say, a bird's eye view of how its components might be arranged for your greater usefulness and ease of comprehension. This vehicle gives you a flexible method for separating, as it were, the woods from the trees."(1)

Clearly some possibilities of this system could be used to explore the concept structures resulting from this project.

RELATIONSHIP TO ARTIFICIAL INTELLIGENCE PROJECTS

In considering the possibility of coding definitions of concepts, propositions and like entities, it is important to benefit as much as possible from related work on artificial intelligence, and possibly pattern recognition. Artificial intelligence projects to simulate human personality or belief systems have had to develop methods and computer techniques which can handle and interrelate entities such as concepts and propositions. Clearly the object of such projects is not attained once an inventory of entities can be examined, even if it is highly structured in the form of a thesaurus. It is therefore interesting to look at both the techniques used to handle concepts and the types of computer-

(1) Lundgren, Nilo. Toward the decentralized intellectual workshop. Innovation (New York), 1971

See also: Engelbart, D.C. Intellectual implications of multi-access computer networks. Menlo Park, Stanford Research Institute, 1970. (Conference paper). For dialogue implications, see U.S.A. National Academy of Sciences Committee on Scientific and Technical Communication (SATCOM), in 1969, that: "More exciting than retrieval of information from a static store is evolutionary indexing, in which user's additions, modifications, restructuring, and critical commentaries, steadily improve the initial indexing..." National Science Foundation funding of investigation into this approach was recommended.

based interrogations that are then possible (1).

The suggestion that techniques of handling individuals' "beliefs" should have some parallel to a community of scholars' attitudes towards the concepts, propositions, etc., which constitute its territory, may appear somewhat provocative. Does a school of thought constitute a belief system?

T.S. Kuhn (2) uses the terms "belief", "metaphysic", "commitment", and "conversion" in connection with a scientific community's attitude towards a paradigm and paradigm change. It might be useful for disciplines to examine their own conceptual structures in the same way as an aid to the development of the discipline. It could be particularly important as a means of highlighting tensions within the conceptual structures which lead up to Kuhn's paradigmatic changes.

This approach suggests a number of stages of sophistication in the possible development of this project.

1. A static inventory of concepts and propositions.
2. A static network of interrelated concepts and propositions
3. "Activation" of propositions as rules governing the relationships between entities
4. Treatment of a school of thought as a belief system
5. Extension to natural language interaction

On this last point, it may be possible to allow a (non-computer-oriented) specialist in a particular field to "dialogue" with the concept data base to permit him to discover and indicate where he differs from its contents and what new he thinks should be included(3). This approach might be a useful method of getting around the behavioural problems associated with the power position of official classifiers in committees.

It may eventually be possible to have many such people interacting in natural language with the data base via terminals to facilitate communication (e.g., at a special seminar].

(1) L. Terler, H. Enea and K.M. Colby. "A directed graph representation for computer simulation of belief systems." Mathematical Biosciences, 2, 1/2, Feb. 68, 19-40

K.M. Colby, L. Tesler, H. Enea. "Experiments with a Search Algorithm on the Data Base of a Human Belief Structure." Stanford University, Artificial Intelligence Project, 1969, (Memo AI-94).

John C. Loehlin. Computer Models of Personality. New York, Random House, 1968

K.M. Colby and D.C. Smith. "Dialogue Between Humans and an Artificial Belief System". Stanford University, Artificial Intelligence Project, 1969. (Memo AI-97)

(2) T.S. Kuhn. The Structure of Scientific Revolutions. Chicago, University of Chicago Press, 1962.

(3) See: K.M. Colby and H. Enea. Heuristic method for computer understanding of natural language in context-restricted on-time dialogue. Mathematical Biosciences, 1,1-25, 1967.

OTHER INITIATIVES

1. Concept coding schemes

There have been many attempts at isolating and classifying elements of meaning at the root of complex concepts. De Grolier(1] notes that methods and the need for them have been regularly discovered and rediscovered since the time of Leibniz or even earlier. He then states: "We draw attention to these 'anteriorities', not in order to underrate the work performed by the various researchers or teams of researchers - who, in most cases., truly believed that they had discovered a 'new method' - but to persuade them, rather than to advocate unilaterally any one 'exclusive' process, to agree that they are all engaged in work on common basic principles, whatever may be the differences (at times very minor] in the coding method or the particular type of machine, adopted."

De Grolier has summarized the work on classification around the world but only a few initiatives seem to be directly related to this project. Usually the work has been directed towards solving a classification problem in some particular field, which strongly influences the design of the scheme. The following, noted by de Grolier, are of more direct relevance:

1.1 Perry and Kent (Western Reserve University]

Developed a coding method for the field of metallurgy based on 'semantic analysis' of complex terms into 'individual terms'. 30,000 terms were assembled from a variety of sources. The notation is however very cumbersome.

1.2 S.M. Newman (U.S. Patent Office]

A "vast attempt at defining or redefining concepts, which could perhaps be entitled - to paraphrase a famous title - 'In search of lost simplicity': to discover or rediscover non-equivocal terms beyond the complications of natural language, which 'unfortunately' does not have "uniform or logical rules for the denomination of devices or things". In effect this is an attempt at creating a metalanguage - but again results in a cumbersome notation.

1.3 C.G. Smith (U.S. Patent Office]

Suggested a system which would isolate "ultimate concepts.... required in the definition of more specific concepts.... There is a basic layer of concepts which do not require definition. It is the use of such elemental concepts which is contemplated in the present system.... A fundamental feature is to seek beneath composite words the basic organization of elemental concepts which they represent, and to develop the essential combination for the definition of these words."(2] This was conceived mainly for patentable contrivances on the US Patent Office Interrelated Logic Accumulating Scanner. It does however permit chains of related concepts to be handled.

1.4 Cordonnier

Worked on methods "to symbolize the elementary points of view of the

(1] Eric de Grolier. A Study of General Categories applicable to Classification and Coding in Documentation. Paris, UNESCO, 1963.

(2) C.G. Smith. Descriptive documentation, International Conference on Scientific Information, 1958] Proceedings. Washington, National Academy of Sciences, 1959, p. 1103.

classification of ideas and...to study the grouping of these symbols in order to obtain composite symbols representing the structure of complex concepts". He also suggests that "intuition permits the representation in an intellectual space of a logical figure, to n dimensions, a synthesis of the relationships between a group of ideas into the different classes which arrange them naturally according to the various possible individual viewpoints".

1.5 M.E. Stevens

Worked on use of computers to handle interrelationships between terms and to 'define', by supplying the generic and descriptive terms related to the term of which the definition is sought; 'develop', by furnishing specific examples of a generic term; 'localize', by indicating the place which can be associated with the proposed concept; 'match', by comparing several proposed terms together, in order to find a 'common point' making it possible to relate to these terms another term possessing the same characteristic; and carry out other logical operations. (1)

2. The ADMINS system

Work has been in progress for some years at the M.I.T. Center for International Studies on the development of very general systems for time shared computer data management(2). An item of data is treated as a sequence of categories of information in n-adic relations applied to a specific entity. N-adic data descriptions for social science propositional inventories are noted as being quite complicated, e.g. 'violence' is 'power' over 'power' over 'well-being'.

The ADMINS system makes use of a "calculus of relations" for stating the derivation of a new relation that draws on those already existing, and which yields a new relational record between particular entities. It is in the structuring of the programming language around the relational record and in achieving intimate interaction with many storage levels that this system differs from most procedure languages.

3. Citation indexing

The citation indexing method described by Eugene Garfield (and implemented in the form of the Science Citation Index and the recently initiated Social Science Citation Index] is of great interest to this project if the focus

CD Mary E. Stevens. A machine model of recall. Paris, UNESCO, NS/ICIP/J.5.4, 1959. See also: T. Kilburn, R.L. Grimdsdale and F.H. Summer. Experiments in machine learning and thinking. Paris, UNESCO, NS/ICIP/5/6/15, 1959.

(2) Stuart D. McIntosh and D.M. Griffel. The requirements for a computer-based information, system. M.I.T., Center for International Studies, 1968, (c/68-14c), 82 p. - Computers and categorization (Paper presented to the Classification Research Conference, Bangalore, 1969). M.I.T., Center for International Studies, 1969 (C/69-283, 41 p.

on documents can be replaced by a focus on concepts(1). The traditional philosophy of classification system design implies that individual entities (usually documents) can be treated as though they were independent of one another. This basic fallacy not only results in the loss of important informational links, but it is basically inefficient- Little or no effort is made to establish a possible relationship between the entity being classified and the entities already classified. There are exceptions to this rule, but generally the building-block development of human knowledge is not perceptibly reflected in traditional classification systems. In conventional word indexing systems, the indexers cannot afford the time to establish linkages between concepts.

Each addition to the body of knowledge is treated as one of a series of independent events, like molecules of a gas. But the literature is not an "ideal gas" - the molecules interact. Similarly, the body of knowledge, partly embodied in the literature, is composed of highly interrelated elements. It is a heavily cross-linked network. The clearly-visible linkages are those ordinarily provided by authors in the form of explicit citations. Less clearly seen are implicit references as in eponyms and neologisms. Almost invisible linkages exist in the natural language expressions which obscure the relationships, especially to an unskilled observer. Conventional bibliography is essentially a simple listing or inventory of publications which disregards most of the interrelationships between the items in the inventory. In contrast, citation indexing integrates this necessary and useful listing in a huge graph or network. In this graph, each entity (in this case documents) is a node or vertex in a huge multi-dimensional network. By analogy, this model of the literature (which Garfield considers to be equivalent to man's knowledge) is like a large road map in which the cities and towns share varying degrees of connectivity. Even the smallest hamlets are nodes on the citation map of science.

Garfield refers to previous work of his on this type of historical map(2). The powerful technique illustrated by Figure 7 is reproduced from one of his papers (1). Since each document is an "event "and bears a date, a graphical history may be displayed, but with the important advantage of being able to show the interrelationships among events. This is a legitimate starting point for the historian.

There is clearly no technical obstacle to handling conceptual entities in the same manner as documents. This would clearly be of value to both the historical and educational model types.

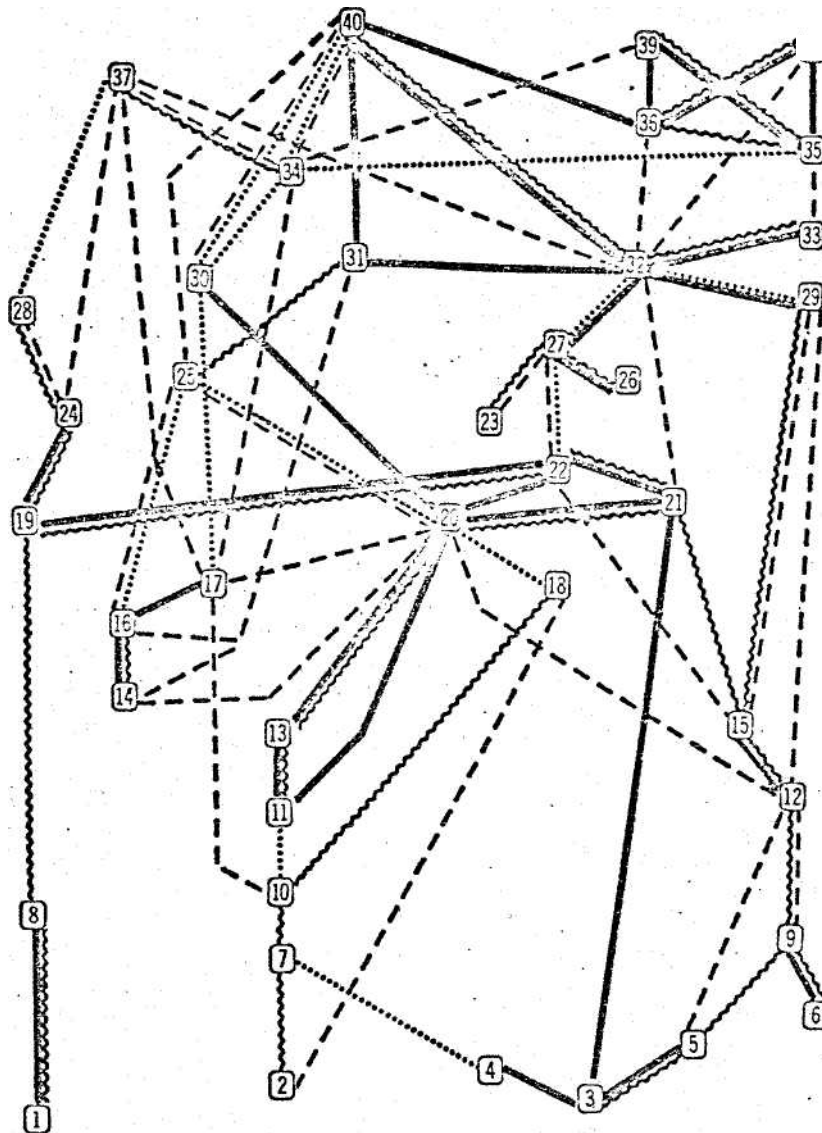
Garfield himself refers to the possibility of having such graphs displayed directly onto a computer-controlled TV screen or plotted onto graph paper by a plotting device. Computers currently plot such graphs on standard line printers as output from the commonly-used PERT programs.

(1) Eugene Garfield. "Primordial concepts, citation indexing and historic-bibliography." Journal of Library History, No. 2(3), 235-249 (1967). see also: Eugene Garfield. "Science Citation Index; a new dimension to indexing." Science , 144, 649-654, (1964)

(23) E. Garfield. Citation indexing: a natural science literature retrieval system for the social sciences. American Behavioral Scientist, 7 (10) 58-61 (1964).

1960
1958
1930
1940
1930
1900
1830

Figure 7



Network diagram for history of DNA based on Asioov's book, *The Genetic Code*. Composite of six network diagrams as reported in E. Garfield, I.H. Sher, and R.J. Terpie, *The Use of Citation Data in Writing the History of Science* (Philadelphia: Institute for Scientific Information, 40 pages).

	NODE	DATE	NAME	NODE	DATE	NAME
	1	1820	Braconnot	21	1947	NAME
	2	1868	Mendel	22	1950	Chargaff
	3	1871	Meischer	23	1950-51	Chargaff
	4	1879	Flcaing	24	1951-53	Pauling and Corey
	5	1886	Kossel	25	1952	Sanger
Direct citation connections	6	1891	Fischer and Piloty	26	1953	Hershey and Chase
	7	1900	DeVries	27	1953	Wilkins
Indirect citation connections	8	1907	Fischer	28	1953	Watson and Crick
	9	1909	Levene and Jacobs	29	1955	DuVigneaud Todd
Asioov's classified historical connections	10	1926	Huller	30	1954-56	Pabce
	11	1928	Griffith	31	1938-57	Fraenkel-Cortrat
Asioov's implied historical connections	12	1929	Levene, Mori and London	32	1955-56	Ochoa Kornberg
	13	1932	Ailoway	33	1956-57	Hoagland Jacob
	14	1935	Stanley	34	1957-58	and Monod
	15	1935	Levene and Tipsoa	35	1957-58	ilurwitz Dintzis
	16	1936-37	Bawden and Piric	36	1960-61	Novelli
	17	1938-39	Cspperrson and Schultz	37	1960	Allfrey and Mirsky
	18	1941	Beadle and Tatum	38	1961	Nirenfaerg and Matthaei
	19	1943-44	Martin and Syngé	39	1961-62	
	20	1944	Avery, MacLeod and McCarty	40	1962	
					1961-62	

Garfield is only concerned with the time or historical dimension as a means of sequencing entities, and only with the citation relationship between such entities. There is no reason, however, why other dimensions and relationships should not be used: geographical, educational, logical, etc., corresponding in fact to more of the model-types listed in an earlier heading.

4. Subject Classification Schemes

There are a wide variety of subject classification schemes for document handling. The Universal Decimal Classification and Dewey systems have become widely used but many other systems exist for specialized subject areas.

The most recent international review of these schemes in the UNISIST Study of the feasibility of a world science information system has this comment to make:

"Librarians and information specialists would generally agree that a world-wide scheme of subject categorization is needed to facilitate document and information exchanges...Opinions differ, however, when it comes to deciding which scheme best suits the purpose. Several encyclopedic classifications are in competition - the Dewey Decimal Classification, the List of Subject Headings used in the Library of Congress, the Colon Classification, the Universal Decimal Classification, etc. -and although the last named has benefitted from extensive international support through FID, it is by no means the unique candidate for worldwide recognition as the standard subject category list. Its advantages and shortcomings were examined by the UNISIST Working Group on Research Needs in Documentation, who came to a twofold conclusion: (a) organizational and technical measures could be taken to obviate the managerial drawbacks of UDC, e.g. slow revision procedure, infrequent re-editions, etc.; (b) on the other hand, no clear answer could be given to the more controversial question of overall or local inadequacy, as regards the content and structures of UDC divisions...further studies and experiments are required to assess the potential value of UDC in its present state, as the unique world list of subject headings for broad categorization, or "shallow" indexing of documents." (p.95]

As the UNISIST extract above acknowledges, UDC is one amongst many classification 'schemes which are in competition. * The tendency for different classifying groups to favour different category breakdowns should be contained and facilitated within an information system and not left to deteriorate into sordid squabbles which do not recognize the value -to knowledge advance of alternative views, and a continuing effort at reconceptualization, restructuring and redefinition of knowledge.

Also of interest is the UN/OECD Aligned List of Descriptors which has now been developed into a "macrothesaurus". This is primarily oriented around mission-focused topics which emerge in the work of the major intergovernmental agencies concerned with economic and social development. From the perspective of this proposal, the following operations have been blurred together:

- entities are labelled by terms
- terms have to be classified into semantic fields to be incorporated
- terms have to be translated and agreed as terms to avoid language dependence

When terms are dropped, the reverse procedure affecting the structure of the list must be followed. Each of these steps involves operational and intellectual difficulties which tend to slow down and resist modification. In addition, the List makes great efforts to be flexible by being term-oriented. To do this it has had to avoid hierarchical classification of any depth. This choice is not in the interests of those users who need a "deep" classification structure.

Originally, (1967-68) it was intended that UNISIST should cover the basic natural sciences but arguments were put forward for the inclusion of technology "or at least some of its branches, especially medicine, agriculture, building and construction". Ultimately, "the position of the ICSU/Unesco Central Committee was that UNISIST should devote its primary effort to the basic sciences...and at the same time be sympathetic to a progressive inclusion of the applied and engineering sciences - and eventually the social sciences - on an equal footing with the former" (UNISIST Report, p. 135-6). No time scale was given.

The special problems of social sciences are ignored in this vague intention to broaden UNISIST. Whilst the latter may prove to be a dramatic success in the field of the natural sciences, it is questionable whether the same techniques can be successfully applied to the social sciences without doing violence to the process by which the latter develop.

In the natural sciences, invariants in the objective world are represented by signs which can in most cases be directly and unambiguously attached to the object in question, to the satisfaction of the natural science community. The sign for the object and the conceptualization of it are intimately and unambiguously related. Another sign in another language may be used but the rules of transformation are clear (the natural language verbiage is another matter, but is less significant). It is a case of "one sign, one concept, one object". It is therefore possible to infer that knowledge transfer tends to accompany information transfer. (This inference may however be very dangerous in the case of non-Indo-European language users, for whom the "objective" nature of the world may appear less significant). But any extension of the world science information system, as it is conceived, to the social sciences would only be of superficial significance if the above distinctions were not reflected in the design of the system. This is because in the social sciences, most of the debate concerns the relation between perceptual invariants detected (by the consensus of a group), signs (selected by the group) and the associated conceptual meaning - as has been recently pointed out by Jean Piaget (1):

"All the social and human sciences are more or less closely concerned, in their diachronic aspects, with the development of knowledge (as a subject)...The foregoing considerations show that the human sciences, in so far as they necessarily include in their field of study the subject of knowledge - the source of the logical and mathematical structures on which they depend - do not merely maintain a set of interdisciplinary relations between one another...but are part of an extensive circuit or

(1) Jean Piaget. General problems of interdisciplinary research and common mechanisms. In: Unesco. Main trends of research in the social and human sciences. Paris, Unesco, vol.1, 1970, pp. 467-528.

network that really covers all the sciences...It was essential to recall this so as to be able to shape our conclusions in such a way that they might succeed in revealing the true significance of interdisciplinary relations.

"For their significance far exceeds that of a mere tool for facilitating work, which is all they would amount to if used solely in a common exploration of the boundaries of knowledge. This way of viewing collaboration between specialists in different branches of knowledge would be the only possible one if we admitted a thesis, to which far too many research workers still unwittingly cling - that the frontiers of each branch of knowledge are fixed once and for all, and that they will inevitably remain so in the future. But the main object of a work such as this...is to push back the frontiers horizontally and to challenge them transversally. The true object of interdisciplinary research, therefore, is to reshape or reorganize the fields of knowledge, by means of exchanges which are in fact constructive recombinations." [p. 521-524, emphasis added]

The natural sciences are therefore primarily interested in the debate on the, usually tangible, content of categories (which are considered to be relatively permanent), and the dynamic lies, in subdividing the categories and discovering relationships between their content. Whereas the social sciences, unable to latch onto an unambiguous content, are primarily interested in the categories themselves and their interrelationships, and the dynamic lies in reformulating, reshaping, and regrouping the system of categories in an effort to get closer to the content(1). It is clear that the natural sciences could easily adjust to an arbitrary permanent category hierarchy, whereas the social sciences would be straight-jacketed and ill-served by any such system.

Perhaps the clearest example of the need for a concept-or knowledge-oriented approach in the case of the social sciences (as opposed to a subject/descriptor approach) is given by the confusion of meanings associated with the concept "democracy". Few people know that Unesco arranged an expert meeting to clarify its meaning. The meeting concluded that at least thirty distinct meanings were required and in use(2). The report was withdrawn from circulation for political reasons - it is political dynamite. It means that in most international debates (in which the word is a vital element of the consensus of interest and common goal on which the discussion is founded) participants are simply talking past one another, and resolutions containing the word are of questionable significance.' In fact, the multiplicity of interpretations implicit in term-oriented discussions and report production may be considered a direct stimulus to the production of further reports giving clarifying or alternative interpretations - thus further clogging document systems.

5. Concept Dictionaries

The outstanding importance of dictionaries in the modern world explains why some lexicographers are dissatisfied with the mechanical method of arranging

(1) Both natural and social science have conceptual parsimony as a criterion, whereas the, "sciences humaines" are interested in multiplying the number of possible concepts and increasing their variety.

C2] Mentioned by F.A. Casadio, Director, Societa Italiana per l'Organizzazione Internazionale.

words in alphabetical order, and would prefer to classify them according to the concepts which they express. One would be mistaken in believing that this is a recent trend, since one finds tentative systematic vocabularies at Babylon in the third millennium before Christ.

It would obviously be a great convenience if conceptual dictionaries of different languages, periods, or single authors could conform to the same general pattern so that they could be readily compared with one another. To this end one would require a conceptual framework so comprehensive and yet so elastic that the most diverse languages and the most idiosyncratic writers would fit smoothly into it. Such a broad classification of concepts was put forward by R. Hallig and W. von Wartburg in 1952.

The German research on "semantic fields", which later inspired Georges Matoré, *La méthode en lexicologie, domaine français*, [Method in lexicology in the field of the French language], Paris, Didier, 1953; offers (p.70-43 a diagram of a "comprehensive classification of lexicon facts" different from that of Hallig and Wartburg, and, moreover, less satisfactory. At the eighth International Congress of Linguists (Oslo, 1957) there was a (rather disappointing] discussion on the subject "To what extent can meaning be said to. be structured? (p. 636-704 of the Proceedings).

Needless to say, the Hallig-Wartburg system is only one of various possible ways in which concepts could be classified; the aim was not so much to devise an ideal scheme as to have a unique basis for specific investigations. If this idea were to be widely adopted, a series of coordinated research projects could be planned with sufficient flexibility to adapt the scheme to the material examined, and yet with enough common ground to make the results comparable.

For a more detailed review of initiatives in this area, see Ullman (1) and de Grolier(2).

#

6. "World Problems" Identification

The author is currently engaged in a project co-sponsored by the Union of International Associations, Mankind 2000 and the Center for Integrative Studies. This is an attempt to identify, "register" and describe worldwide problems with a view to the publication of a Yearbook of World Problems(3). (Work to date has established that there might be some 2000-5000.) The approach is similar in philosophy to that proposed here for concepts. Classification of problems is seen as a second and distinct phase. -A crude model is being used to facilitate data collection. - Two other models will be used to plot problem interrelationships. It is hoped to be able to map and plot problem networks.

(1) S. Ullmann. *Semantics: an introduction to the science of meaning*. Oxford, Blackwell, p. 254-5.

(2) E. de Grolier. *A Study of General Categories Applicable to Classification and Coding in Documentation*. Paris, Unesco, 1962, p. 226-228 (Note 89).

(3) To be a sister volume to, and cross-reference, the UIA's Yearbook of International Organizations, which is now produced via computer permitting access to data for research purposes.

LANGUAGE AND TRANSLATION PROBLEMS 1.

Absence of a lingua franca

It would be optimistic to expect wide acceptance of the system if it was based on one language only. The UNISIST Study notes (pp. 72-73) that:

English now accounts for about 40% of the world literature, regularly yielding (as are French and German) to the rising group of "Eastern" languages, e.g. Slavic, Chinese and Japanese.

No one can predict what the situation will be twenty or fifty years ahead, nor does anyone possess reliable data on the present use of foreign language materials in the scientific community.

The position of English as a lingua franca of science is contested by some governments either to consolidate a new country via a national language or in the belief that language can be artificially maintained as a vehicle of a culture.

The chances of securing international acceptance of English as the standard language of science are, in present circumstances, very poor.

2. Language preferences

Apart from these aspects, there is the extremely serious problem that social scientists in one language group tend to either ignore foreign language material or find it "less relevant" to their particular concerns. This is particularly significant across the English, French, German divide. Concepts given in foreign languages may be difficult to comprehend if one is less than completely at home with the language in question. An unconscious hostility to concepts expressed in foreign languages may even build up.

A recent study of 1000 social science research information users in Great Britain has just been completed (1). It shows that 18% of the sample read English only, 75% read French, and 27% read German. Of those who said they were able to read a foreign language, only one-third regularly scan literature in that language. There is even a reluctance to follow up articles in another language.

It was also noted that 22% make no use of abstracts or indexes, 35% never use bibliographies, 22% do not use library catalogues, and 48% do not consult the librarian.

3. Language group incompatibilities

There is also the possibility that a concept may first be expressed or may only be expressible, in a given foreign language. It would be an advantage to be able to file it as such and worry about the translation afterwards. The author who has done much to emphasize the difficult-to-comprehend contrasts between meanings in the standard Indo-European languages and those in other language groups is Benjamin Lee Whorf (2). He suggests that language becomes

(1) Maurice Line [Ed], Information Requirements of Researchers in the Social Sciences. Bath University, 1971, 2 vols.

(2) B.L. Whorf. Language, Thought, and Reality. New York, Wiley, 1958, 278 p.

a classification and organization of experience in its own right. As such each may be significantly different from the other and may structure the forms and categories by which the individual not only communicates but also analyzes nature, perceives or neglects particular phenomena or relationships, and constructs his model of the world.

A striking example of the possible differences is given by Marshall Walker in discussing the social factors which affect scientific models:

"The language of the Wintu Indians of California seems to indicate a way of thinking quite different from our own. Imagine the surface of a table with a book lying on it. The remainder of the surface is bare. In English one describes the situation by saying "The book is on the table". In Wintu one says, "The table bumps". The English phrase has already committed the speaker to an entire analytical philosophy of the situation: (1) there are two objects; (2) there is a polarity such that one object is above the other; (3) there is an implication that the book is supported by the table. None of this analysis is present in the Wintu sentence, which is purely topological....The scientist who wishes to be as objective as possible in his study of the external world will try to free himself from the possible constraints of his own language."(1).

Such languages may not have parts of speech or separate subject and predicate. In Indian Languages such as Nootke and Hopi events as a whole are signified. Instead of "a light flashed" or "it flashed", Hopi uses a single term, "flash", to signify that a happening has occurred. There is thus no distinction between tenses, for the Hopi has no general notion or intuition of time as a smooth flowing continuum in which everything in the universe proceeds at an equal rate, out of a future, through a present, into a past. Marshall Walker also notes (p.103-43:

"The student of science also has a vital need for comparative linguistics in order to acquire experience in the isolation of concepts from their language matrix. The usual language departments of a university are not much help for this type of study....There is need for a course for undergraduates (not language majors) which is designed to illustrate the expression of concepts by different language families. Pending the arrival of such courses the student of science will have to do it himself as best he can."

David Bohm, a theoretical physicist interested in Piaget's and Gibson's work on the problems of perception, gives detailed arguments against permanence of "entities" and concludes(2):

"it is clear that both in common experience and in scientific investigations, the objects, entities, substances, etc., that we actually experience, perceive, or observe, have always (thus far³ shown themselves to be only relatively invariant in their properties, this relative invariance having often been mistaken for absolute permanence"
(p. 14)

(1) Marshall Walker. *The Nature of Scientific Thought*. Prentice-Hall, 1963, p. 103.

(2) David Bohm. *The Special Theory of Relativity*. N.Y., Benjamin, 1965.

"It is evident then that by considering entities and structures as relatively invariant, with an as-yet-unknown domain of invariance, we avoid making unnecessary and unprovable assumptions concerning their absolute invariance. Such a procedure has enormous advantages in research, because one of the main sources of difficulty in the development of new concepts - not only in physics but also in the whole of science - has been the tendency to hold onto old concepts beyond their domain of validity." (p. 121-23)

4. . Problems of translation

It may astonish many people to know that contemporary linguistics has concluded that translation between languages is theoretically impossible, Chomsky notes (p.202):

"In fact, although there is much reason to believe that languages are to a significant extent cast in the same mold, there is little reason to suppose that reasonable procedures (not involving extra-linguistic information) of translation are in general possible."

Georges Mounin, who notes the same conclusion, has summarized the theoretical difficulties prior to considering why, how, and within what limits the practical operation of translations is relatively possible (1).

Some of the difficulties he notes argue against any attempt to force this project into a unilingual mode.

- certain languages have highly developed terminologies in areas where there are few Indo-European equivalents (e.g. the Pyallup Indians and "salmon"; the Eskimos and "snow" (30 terms), some African languages and "palm trees", the Argentine gauchos and "horse colouring" (200). There is little value in attempting a definitive translation when no exact equivalent exists.
- the situation becomes more complex when dealing with socio-cultural terms, e.g. how can "brother" and "sister" be translated into Maya when that language only has terms for "younger brother" or "older brother"(2). Much closer to the concerns of this project is the simple problem of translating "people's capitalism" into French(3).
- another excellent example, noted by Colin Cherry(4) is that whilst there is no difficulty in translating the colour "red" into and from Russian, the associations in the two languages are very different. In English: blood red, red in tooth and claw, red with anger, red light district, etc. In Russian the translation of "red" is synonymous with "beautiful" and has associations equivalent to the English "golden" - hence "Red Square" and the "Red Army" should be meaningfully translated as the "Golden Square" and the "Golden Army". (How

CD Georges Mounin. Les problemes theoretiques de la traduction. Paris Gallimard, 1963.

(2) A special issue of the ETC (Institute of General Semantics), 15,2, March 1958 is entirely devoted to interpretation and intercultural communication. It gives many examples of this sort of problem.

(3) Georges Mounin, op.cit. p. 67-68

(4) Colin Cherry. World Communication; threat or promise? London, Wiley, 1971, p.16

much has international tension been aggravated and reinforced by this simple error?) Similarly, in Chinese, "red" is primarily associated with "joy", "prosperity", "luck", and "happiness"(1). Thus greeting cards, invitations, decorations, etc., are usually in red. (To what extent have the positive associations of the colour in the two cultures influenced the marked success of socialism there, compared to that in Anglo-Saxon culture, where it has more negative association?)

5. Administrative delays

If the attempt is made to translate every theoretical formulation into English, before filing, there will be a hold-up similar to that associated with the modelling activity. There is also bound to be disagreement as to the adequacy of translations. It may be preferable therefore to conceive of a Translation Phase in parallel with the filing, modelling, and term allocation phases, and to give priorities to the translation of given terms according to need.

DOCUMENTATION OR KNOWLEDGE?

1. The Documentation Problem

It is the stated goal of the UNISIST World Science Information System to facilitate the "unimpeded exchange of published or publishable scientific information and data amongst scientists in all countries". Its concern is therefore with the extremely large number of documents and not with the relatively limited number of original conceptual entities formulated therein. Unfortunately, the UNISIST Study does not distinguish between documentation, information and knowledge(2).

Briefly, documents pose a physical handling, transfer and filing problem (which may be eased by reproduction at a distance). Information consists of signs which can be read, transferred, manipulated and filed electronically. They function as symbols of units of human knowledge, but only during the short-duration process of being read for meaning. Knowledge transfer depends on the ability of the momentary psychological system "sign and reader" to generate an unambiguous, coherent and consistent meaning in the mind of the reader, and conversely to convert a distinct meaning or concept into a suitable sign which can be interpreted with equal ease by another reader. Information, in the form of signs, can be read without resulting in the transfer of knowledge and particularly of the knowledge intended [e.g. undecipherable hieroglyphic writing can be "read" without knowledge transfer).

The Study does not recognize that the period covered by the proposed system is one in which increasingly, it is almost impossible for the decision-maker or researcher to determine what information from which discipline is

(1) I am grateful to Mr Thai Wo Tsan for this information.

(2) See UNISIST Study Report, op.cit., p.1, p.20, p.103, p.115, p.148, p.152 (This point is examined in more detail in COCTA Working Paper No. 3, p.65).

"relevant" (1). If he attempts to order all the relevant documents (or even subscribes to the appropriate abstracting service), the purchase or transport costs will be prohibitive (except to a small elite); if he waits for all the relevant information, it will be too late for him to make a useful decision(2); if he gets all the relevant information in the form it currently takes, he will have neither the time, the training, nor the inclination to read it all; and if he reads and comprehends it all, he will not have the time or the ability to convey his understanding to those whose support he must obtain to carry a vote on the matter or, ultimately, to the man in the street.

"Consider this dilemma: while our technological abilities to generate and disseminate potentially useful data have increased manyfold in the past few years, man's physical capacity to register and to process potentially informative data has probably increased very little, if indeed at all. The sheer volume of data that crosses the typical executive's desk today should serve to spotlight the inadequacies of the education and development of our acquisition strategies and practices. But no gain in ability could offset the widening gap between the exponentially-increasing quantity of data available for consumption and man's very limited capacity for acquiring and processing useful information."(3)

It is questionable, in view of present trends, whether knowledge transfer can continue to be effectively accomplished primarily via document transfer. The United Nations is potentially the most significant institution in existence and is at a vital nexus of multidisciplinary, international knowledge transfer - which it currently accomplishes via documents(4). And yet it has a documentation problem (which in a sense is equivalent to that of many, if not most, other large organizations and disciplines):

"This issue has been repeatedly recognized by the General Assembly, the Economic and Social Council, the Joint Inspection Unit and nearly a dozen of other UN bodies as one which directly affects the functioning of the UN. Suffice it here to note that in 1970, the UN, both in New York and Geneva, produced nearly a million page documentation in all languages. The massive volume of documentation produced by the UN prompted a former President of the General Assembly, Mr Lester B. Pearson of Canada, to remark that "the United Nations is drowning in its own words and suffocating in its own documentation." The Joint Inspection Unit stated recently in its report submitted to the present General

-
- (1) "...how is a practitioner of any one discipline to know in a particular case if another discipline is better equipped to handle the problem than is he? It would be rare indeed if a representative of any one of these disciplines did not feel that his approach to a particular organizational problem would be very fruitful, if not the most fruitful..." (R.L. Ackoff, Systems, organizations, and interdisciplinary research.)
- (2) The author recently had to wait seven months for an in-print ordered publication. Its title: Foundations of Access to Knowledge. Syracuse University Press, 1968.
- (3) Lee Thayer. Communication and communication systems; in organization, management, and interpersonal relations, Homewood, Irvin, 1968, p.202.
- (4) UNITAR/EUR 3/2, 1971, p.2. "Only recently the Secretary-General of the United Nations affirmed that the Organization's most important working tools were documents. Thus the main medium for conveying information consists of documents."

Assembly session that "the inspectors do not hesitate to say that the point of saturation has now been reached and indeed overstepped." (1)

The last quote in fact continues with the significant phrase "and that the law of diminishing returns is taking over...Beyond strictly financial considerations, therefore...the future usefulness of the Organization may well hinge on its ability and determination to set once and for all, and strictly enforce a reasonable but drastically reduced ceiling to the volume of documentation its various bodies call for and its services produce" (2).

One is not exposed to alternative hierarchies of conceptual nexuses linked directly or indirectly to more distant nexuses from which relevant knowledge may be obtained. (There are no "heights" in documentation systems - the general is filed with the particular, cf. the treatment of documents with an interdisciplinary emphasis.) The potential value of a knowledge-oriented information system as an active stimulus for creative social change and problem-solving may even be directly proportional to its ability to draw attention to the existence of established relationships of low probability (i.e. low entropy) between concept nexuses. This is not a criterion of document information systems where the emphasis is - for cost reasons - on facilitating access to those documents which are most probably relevant in terms of demand frequency.

Shuffling documents and signs might facilitate the transfer of meaning and knowledge between those who could identify the representative of the group for whom a particular set of meanings could be consistently and unambiguously attached to the signs. But even within that group, advances in knowledge and reconceptualization have to be carefully related to the original set of meanings.. However, making the documents and signs of that group available to other "outside" groups would only introduce "noise" and confusion. A knowledge-oriented information system would be needed to avoid such confusion and facilitate fruitful interaction between different schools of thought within the social sciences.

(1) UNITAR. The Interest of the United Nations Institute for Training and Research in the question of United Nations documentation. Geneva, UNITAR/Eur 3/1, 1971, p.1.

UN Document A/7576, 25 July 1969, para. 2, shows that document production by New York HQ increased by 50% from 1964 to 1967, to 600 million page-units. This does not include production of the regional or Geneva offices or specialized agencies. A recent UNITAR document (UNITAR/Eur/3/2 notes that there will probably be one million journals in 30 years time. Currently it is estimated that about 2000 books (i.e. 1 million pages) are printed every minute throughout each day.

(23 UN Document A/8319, 2 June 1971 (or JIU/REP/71/4). But stemming the generation of new knowledge in developed countries, is about as feasible as lowering the birth rate in developing countries. To severely reduce one means of storing and disseminating such knowledge, without seeking a more appropriate complementary medium, could only be counter-productive and unsatisfactory.

2. Knowledge-Representation

The ideal "information" system in a given academic field has been sketched out by the U.S. National Academy of Science Committee on Information in the Behavioural Sciences [under the chairmanship of David Easton) as the "computer analogue of the available, intelligent, and informed colleague". There have been many reports on the improvement and integration of information systems and it would be futile and inappropriate to comment on them here. There seems, however, to have been little mention of what might be termed a "knowledge-representation" system.

In parallel columns below, an attempt is made to clarify the implications of this distinction by comparing the functioning of a hypothetical knowledge-oriented system, now technically feasible, with the current approach. The intention is not to imply that the former should replace the latter but rather that the former offers various means of avoiding some of the key problems faced by the latter - the two are however complementary. The distinction is basically between a synthesis or atomisation in the handling of information.

Document/Information System

- . Index tends to be based on simple hierarchy or alphabetic listing of subject, author and title, which can be handled on catalogue cards.
2. Users want documents; the index is a temporary inconvenience to gain access to the document.
3. Access to knowledge via documents . means multiple reproduction and transfer of documents to a variety of libraries where they may or may not be used.

Documentation system is embarrassed when faced with obtaining "ephemeral" or "phantom" material which has not been made commercially available through the few standard channels.

Knowledge-representation System

1. "Index" constitutes a complex network giving a representation of entities and relationships and the dynamics of any points under debate, which can only be handled by multi-dimensional computer programming techniques.
2. Users want access to the "network index" which represents the needed items of knowledge and their relationships; documents are a temporary inconvenience only used if it is necessary to re-examine data and detailed arguments justifying the entities and relationships incorporated. [Document access is a secondary problem for which a documentation system may be used.)
3. Access to knowledge is direct and does not require reproduction and transfer of documents. (Only one copy of the document justifying the amendment need exist on microfiche so that copies need only be prepared when the data and arguments must be re-examined in detail.)

4. See 3.

5. Out-of-date, rejected, low quality, false, old documents are retained in the system and indexed with no index indication of their status.
 6. Only the Knowledge held in the documents physically available is accessible. The index only notes the documents held in the documentation centre in question.
 7. Alternative concepts or contradictory evidence can be conveniently ignored in a document or textbook without too much risk - particularly where the counter argument comes from another discipline or a school of thought publishing in a different language).
 8. Interdisciplinary links are ignored if the author has no interest in them.
 9. Different styles of documents are produced on the same topic for research, education, public information, program management, policy making, etc., purposes. The same material is repeated, with some extensions and some omissions, for each audience. This leads to a "spastic" or "aphasic" response to new situations, by different portions of society.
 5. Out-of-date, rejected, false, etc. entities or relationships may be eliminated from the system by listing them on paper (or other "documents") with the bibliographical source from which they were obtained (i.e. they are available if required but do not clog the system).
 6. All knowledge is on-line, although the supporting documents may not be physically accessible.
- Alternative concepts, relationships or contradicting evidence are immediately forced on one's attention even in the case of relationships linking to other disciplines.
8. Interdisciplinary links are already held in position whether the author wants to ignore them or not.
 9. The entities and relationships entered on the basis of research insights are also used for other purposes. Instead of producing different documents and reprocessing the insights, different "filters" are used in presenting or displaying the entities and relationships to different audiences. In this way, each new research insight is immediately incorporated into each other form of knowledge-representation - each portion of society works from the same data base. (Problems registered by non-research bodies are immediately evident as a challenge to research.)

In this way if an element of knowledge represented cannot be understood, the user merely calls for a new method of representation (of the same knowledge) possibly using isomorphs (or even analogies) from a domain with which he is familiar. (At any point he can move into a programmed learning mode and work from simple representations.)

10. The documentation system does not permit panoramic summary of any permanent representation of knowledge in a particular domain. Each verbal summary extant at a particular moment is under criticism and subject to reserve from different schools of thought within the discipline. In this important respect a document arising from a single group of authors can never contain the totality of views in a domain of knowledge. Only the non-concretized interaction between a succession of documents approximates to it. These invisible qualifiers on any document are a feature of the "collective mentality" of the members of the discipline. The knowledge of the discipline at any moment is very much in (and between) the minds of its members rather than on paper or in a row of books.
- The forum of academic debate is concretized as a scattering of journals and other documents. There is little interaction between the journals but the debate is somewhat summarized in a scattering of abstracts in which the contents index gives some indication of the interventions on related topics.
10. Each entity, link and qualification is indicated in the knowledge representation system. In effect one "layer" of the "collective mentality" of a discipline is rendered visible. Each modification to knowledge in the domain can be entered on an hour-by-hour basis.
- The knowledge representation system constitutes a thinking forum in which the juxtaposition of relevant ideas from all sources is maximized. The researcher is exposed to a pattern of theoretical formulations in the process of being continually improved, and to which he can contribute. A dozen or more specialists in a particular field (the "invisible college" for that topic) can contribute simultaneously to work on ideas being written on one memo pad via electronic dialogue support systems which help them to respond to each other's ideas (even if they are a continent apart) with a rapidity that allows each of them to maintain thinking momentum. Even in such a rapid debate the paternity of each emerging formulation is identified and registered. (This mode of operation should be compared with some discussions between academics interested in the same topic in which progress is frustrated because if someone thinks of a good idea he wants to "publish" it (to gain credit) before contributing to the thinking momentum of his colleagues - this may take months.)
11. Thinking momentum is constantly interrupted when access to new documents, is required. (Long delays, 2-3 months, are normal; 50 months or more from initiation of research to appearance in abstracts).
11. the essence of any new domains of knowledge is always accessible -- all the links and entities are there (delays are measured in seconds).
12. Research is conducted primarily using documents as a stimulus to creativity.
12. Research is conducted primarily using the knowledge-representation structure (i.e. the graphical representation) as a stimulus to creativity.
13. Author has "published" when document is in circulation and "available"; index entries of little significance to author.
13. Author has "published" when the appropriate knowledge structure in the "index" has been modified; incorporation in "index" (through a terminal) is of highest priority for the author.

14. Author's status, credibility, pride and interest are associated with visible documents on library shelves. The documentation problem is aggravated by the "publish or perish" code which governs much of academic life. Unless an academic produces a document he is "invisible" and loses status.
14. Author's status, credibility, pride and interest are associated with the visible entities and links in the graphic representation accessible to all. By switching emphasis to the specific entities and relationships which the academic has formulated, successfully, confirmed or criticized -- his status is determined by the bonds and entities with which he is associated. Each of his contributions is "visible" until it is superseded and is not subject to the vagaries of the documentation system.
15. Each new document must carry a lot of 'verbal packaging to provide a context within which innovative elements are introduced. There is no guarantee that the rephrasing (necessary for status and copyright reasons) of earlier arguments will constitute an improvement.
15. The author need only enter the specific entities of relationships which constitute his innovation. (Since the academic's status is bound up with his specific modifications to the knowledge structure and not the verbalizations held in a document, the problem of adequate verbalization may be handled separately. Hopefully a limited number of skilled verbal presentations, from a minimum number of different perspectives and literary styles, could be constantly updated by professional writers using the best verbal arguments by any appropriate academics where appropriate.)
16. The direction of research is governed in part by shifting patterns of credibility and status. These are merely evident in print but are controlled by an ongoing informal dialogue centred upon the elders of the discipline who legitimate consideration of particular entities and relationships.
16. It is quite evident which issues are currently under debate and the manner in which the demise of a set of entities and relationships will weaken the status of a whole set of dependent elements. Ideally the system would also act as a continually updated voting board for each element, providing an opportunity for members of the profession to indicate their " approval.
17. The key figures in a discipline and the relationship between their spheres of influence are unclear.
17. The "luminaries" in a particular discipline are all visible together with the relationship between their spheres of influence.

SUMMARY OF ADVANTAGES OF THIS PROPOSAL

Most earlier initiatives and proposals examined seem to fall foul on one or more of the following difficulties:

1. The simple and unambiguous administrative task of filing entities is merged into the complex intellectual task of coding and classifying them. This makes the whole project lengthy, costly, and complex.

In this project the identification of entities to be included in a thesaurus and the practical problems of incorporating these entities into an information system are distinguished from the theoretical problems of classifying and interrelating such entities. The first is a relatively fast and unskilled operation and the second is a relatively slow and skilled one.

The technique of identifying the entity within the system by a numerical tag derived from a classification scheme is avoided. The savings in labor associated with this technique are only significant in a system in which all operations are manual. Where computers can be used, the two types of operation can be distinguished in order to save resources, speed up operations and increase the flexibility of reconceptualization of any classification scheme.

2. The classification of theoretical constructs may be associated with an intellectual and material investment in a document physical-location system. This opposes any flexibility or major reconceptualization of relationships between entities.

In this project there is no direct relationship between the classification scheme(s) and the physical problem of locating source documents.

3. The classification scheme may be rigid and "final", based upon a high commitment to a particular set of theoretical assumptions of limited comprehensiveness, and therefore unable to adapt to new types of interrelationships.

In this project both rigid and rapidly evolving classification schemes can be used to interrelate the entities handled.

4. The classification scheme may be exclusive or "inhospitable" and therefore of limited use. • .

In this project both exclusive and hospitable schemes may be used. This gives it a wide range of uses.

5. Some systems are specifically designed with the special problems of a particular field of knowledge in mind. This makes them difficult to use in other areas.

In this project specialized and general overdesigning the information handling system to meet immediately-perceived needs would reduce its usefulness and relevance to others and therefore increase the difficulty of ensuring adequate funds over a long period. [The degree of "hygiene" introduced may be inversely proportional to the utility or relevance of the system to potential users.]

6. Even adequate universal schemes may become viewed as authoritarian and a vehicle for some form of conceptual imperialism. Unfortunately the organization of relations between entities is equated with the imposition of a new set of relations. The organizers are perceived as acquiring power. Exclusive or rigid schemes, once created, are viewed and defended as unique and "universally applicable" by their proposers, thus eliminating any possibility of more comprehensiveness, better-funded, joint efforts.

In this project, every effort has been made to ensure that it does not become associated with particular schools of thought, organizations or personalities who might resent criticism of their perspective and alienate potential collaborators. All such individualism is contained within the model building activity which does not jeopardize other models or the project as a whole.

7. The actual procedures for incorporating new entities into any "approved" list within the system may appear bureaucratic and stultifying unless the system is user-oriented. There is therefore the old problem of minimizing the bureaucratic desire for due process and order and maximizing user participation.

In this project suggestions have been made concerning means of maximizing user participation.

8. The system may be designed with only one type of user in mind, e.g. scholars or students. New systems, which compete for the same resources, then have to be created for other users of the same data.

In this project some consideration has been given to methods of introducing "filters" in conjunction with special models in order to show special relationships between entities in a manner significant to other types of user.

Some of the needs of users not immersed in the Western cultural perspective have also been considered.

9. The notation used to indicate the position of an entity in a classification scheme may be very complex. This may make data handling very difficult.

In this project it is not necessary to use a notation in order to file the entity. Only a simple sequence number is required. A standard notation for use in print, but independent of the organization" of the system, has been suggested.

10. The system may be viewed as a "one-shot" job using all the appropriate specialists. This is the case with some concept directories. Even so, non-participants criticize the position taken by the participants, thus suggesting the need for new projects.

In this project it is not necessary to limit classification to the views of one specialist. A number of competing specialists can participate-together or separately without jeopardizing the ability of the system to adapt and respond to new proposals.

11. Systems may be slow (up to decades) in responding to proposals for change, to the point of acting as a constraint on innovation to those dependent upon them.

In this project, modifications and alternatives can be handled without difficulty.

12. A system proposal may raise problems of standardization for purposes of handling bibliographical or other data. The system design may then become a pawn in the debate between the different schools of standardization and information handling.

In this project there are no features which could become a major issue in the ongoing debate, since it is not a conventional documentation system and does not have major bibliographical concerns.

13. A system proposal may constitute a threat to other systems competing for the same resources -- particularly if major changes are proposed for existing systems.

This project does not appear to compete with other systems. It can be considered complementary to some documentation systems.

14. A system may demand, or be designed for, complex computer systems to the point of being unusable in less-richly-endowed environments.

This project is based on a very simple filing system for entities and relationships between them. The resulting file may however then be subjected to analyses of varying power depending on the computer environment available.

15. A system design may raise fundamental theoretical issues, and therefore alienate important potential supporters.

In this project the accent is on providing a simple technique for filing entities and relationships in a way which permits a number of general analytical and display techniques to be used. Every effort has been made to avoid giving a final and exclusive definition of what is incorporated. Such theoretical debates are carefully confined to the activities of modelling groups which are each free to ignore or accept entities and relationships filed by other modelling groups.

NEXT STEP

The next step is to obtain critical comments on the various proposals put forward and to undertake pilot projects in some of the following areas:

-file organization and computer program development or adaptation
operational and logical problems of classification with models
in a few test areas

computer simulation of file movement, modelling activity and
behavioral complications in a decentralized, minimum
organization environment

- computer simulation of different strategies to keep the system "open"
without it becoming uncontrollable

preparation of a graphics demonstration program as a means of
generating further interest and showing the power of this technique.

Exactly how much pilot project activity is required will depend upon the speed with which it is desired that the project as a whole should move forward and the range of interests it is desired that the project should serve. These must be decided.

No comments have been made on the funding required since cost estimation depends on decisions taken for the next stage. The computer programs envisaged for the filing of entities and relationships and generation of lists and thesauri are however fairly simple to prepare and cheap to run. The other major costs would be collection of conceptual entities (unless done voluntarily by a team using existing material), administration (unless incorporated within the budget of some existing institute) and travel costs of those concerned with modelling (unless it was decided to switch immediately to the postal modelling concept outlined).