

The Relevance of INTERCONCEPT for Classification and Indexing*

Vasarhelyi, P.: **The relevance of INTERCONCEPT for classification and indexing.**

In: Intern. Classificat. 7 (1980) No. 1, p. 6–9

Presentation of the ideas and aims of the Unesco Interconcept Project also intended to become a programme of either the Unesco or any other institution interested in the promotion of international understanding by means of conceptual control and analysis in the fields of the social sciences. The article outlines some of the problems yet to be solved, the solution under study, and the organizational matters. It reports also on some of the findings of the Pilot Project.

I. C.

1. What is Interconcept?

Interconcept is intended to be both a system and a programme. The final aim of Interconcept as a long range programme is to promote international understanding by means of conceptual control and analysis in the field of social sciences. Interconcept, as an international information system is supposed to collect, store, process and disseminate information on the concepts attached to key social science terms, to describe and explain these concepts and their significance to various schools of thought and disciplines as well as to monitor the changes that ideas undergo in time and space.

At present, Interconcept is in the pilot project stage initiated and launched by Unesco within the framework of its activities towards the development of social science information. The pilot project is a means of:

- testing various approaches to the collection of conceptual and terminological data;
- establishing the practicability of creating a data base for conceptual and terminological information;
- demonstrating what kinds of products can be prepared using the data base;
- and evaluating the feasibility of establishing a conceptual and terminological information system.

2. Some problems to be solved

The basic problem from which the demand for an international programme and information system on concepts

* Paper presented at the 3rd Annual Conference of the Gesellschaft für Klassifikation eV, Königstein, 5–7 April 1979. Printed in the Proceedings volume of that Konferenz (Klassifikation und Erkenntnis, Vol. II, p. 171–181, "The implications of the Interconcept Project for classification and indexing").

arose is the 'software' of the current social science terminology. (1)

As a matter of fact, social scientists often use the same words for different concepts, but the opposite occurs as well, i.e. they use different words for the same concept. In other terms, two of the fundamental problems that arise in the course of indexing and information retrieval are homography and synonymy.

As to the problem of homography, Prof. Riggs stressed in connection with the pilot project that its source can be found in the procedure followed by social scientists when identifying a new concept that they need to use for theory-building and/or empirical work. They take namely, often a familiar word and "simply assign to it a new meaning. If it were easy to take note of and remember the new meanings, no difficulty would arise. However, the fact is that when scholars use these familiar words it is often not evident to their readers (or listeners) just what their intended meaning is. They, of course, know the word, but its older meanings come to mind, rather than the new one". (2) The receivers of social science simply interpret the terms as best they can, which results in a situation where the understanding of intended messages becomes fuzzy and imprecise.

In the course of the Interconcept pilot project, the analysis of the various concepts for which the same term 'development' is being used, clarified that they can be grouped into three classes: (2)

1. concepts, that are not relevant to social sciences;
2. concepts used by social scientists, but borrowed from other disciplines, or not well-defined, or used in a quasi-scientific way only;
3. concepts that are unquestionably distinctive and necessary in social science investigation, research and application.

Indexers, in the absence of a glossary of social science concepts, find it very difficult, if not impossible, to make the necessary distinctions. Thus, if 'development' is used as a key-word, it would be unfair to criticise indexers who use it as an index term. The result, however, is that indexing may bring together documents dealing with different topics.

If it is ensured by some means that a distinction is made between the above classes and documents which are relevant to the social sciences are selected, the concepts described by the term 'development' will still belong to several groups which Mr. Riggs characterized as follows: (3)

- conceptions of development, considered as a kind of process that enables the social or political system to perform more effectively;
- conceptions of development that focus on the processes by means of which a social system decides what it wants in the light of available alternatives and then carried out its decisions;
- conceptions of development considered as the consequence of a process which led to the successful accomplishment of a course of action;
- conceptions of development that focus on the consequences of the emulation by someone of outcomes previously accomplished by someone else.

Distinction between the above conceptions would be of great help to researchers dealing in most cases with one of them only. In the absence of a glossary of social science

concepts, however, this distinction is practically impossible.

It is not less difficult to identify the documents whose authors use different terms, in spite of the fact that they deal with the same concept. Synonymy creates particularly serious problems because — as Doyle points out — “. . . it causes a group of entries which ought to be under the same heading to be scattered among several widely separated locations in the index”. (5) The result is that, in the field of social sciences, information retrieval based on key words tends to be hopelessly imprecise. As a consequence, users of formal information systems often find that after a good deal of work they obtain materials that are not truly relevant to their interest. In disgust, they may turn to informal systems, relying on friends or acquaintances to procure the information they need.

The obstacles to precise thinking and to retrieval of relevant information caused by the softness of the terminology in the social sciences are admittedly acute within the context of one school of thought and/or one region alone, but they are multiplied exponentially when confronted between different schools of thought, between regions having different socio-economic background and between scientists speaking different languages.

Thus the demand for information on concepts may arise in the following cases:

- a) a scientist (or a group of scientists) needs to know and understand the meaning(s) of a term forming part of a system of concepts within a school of thought which is known to him;
- b) a scientist needs to know and understand the meaning(s) of a term forming part of a system of concepts within a school of thought which is not or not well enough known to him;
- c) a scientist needs to know the equivalent(s) of a foreign language term forming part of a system of concepts within a school of thought which is known to him;

In the first case, the term may be and should be explained within the framework of the system of concepts it belongs to (e.g. a Marxist term should be explained by the means of the vocabulary of Marxism). In the second case, however, understanding can be assured if, and only if, the meaning of the term is explained within the framework of the system of concepts which is known to the scientist (e.g. a non-Marxist term should be explained to a Marxist scientist using either terms having a common meaning for both schools of thought or terms of the Marxism).

In the third case, the scientist needs a literal translation of the term together with a literal translation of its definition.

3. The solution under study

Interconcept is intended — among others — to remedy at least to some extent, the situation created by the softness of social science terms. Its aim is to provide the special class of information; the semantic one, which is now often entirely missing.

With a view to achieving the above, the data base under construction within the framework of Interconcept contains the following types of information on concepts:

- definitions of terms (content analysis is needed here to define the meaning of terms);
- equivalent terms in different languages and/or schools of thought;
- contexts (or explanations) in which a term expressing a concept was born and/or typically used;
- relations between terms, reflecting also the differences in the structure of concept-systems of various schools of thought.

Let us analyse these types of information in some more detail.

(a) The definition of a term in a few cases is already standardized at the international level. It is far more usual, however, that a term has various definitions in various regions and countries. Some definitions are standardized at regional or national levels, others are only harmonized (at some of the above levels), that is, e.g. that they are defined in a terminological vocabulary, widely used internationally, in the region or in the country and have by means of its use a harmonizing character.

The data base may include terms also which have no real definition, but only a description that often does not meet the scientific-logical requirements of a definition.

In order to clarify even more fully the concept behind a term, it might be useful to include in the content of the central data base some extracts which illustrate the use of the term. In some cases, this context will replace a missing definition.

Separate definitions establish the existence of different concepts even though a single term (homograph) is used to refer to several concepts. Consequently, a separate entry is required for every concept definition — even though the same term appears in many entries. Judgements about the importance of a concept will be made by users — obviously many concepts recorded in the archive will not be used, but at least their existence and source become a matter of record.

(b) Relationships. Information about broader, narrower, overlapping and related concepts should be provided on each entry. Since all such relationships will not be apparent when the original entry is made, it is necessary to enable users to provide such information, as comments that can be added to each entry.

(c) Synonyms as equivalents in the same language have to occur under the same key-term. Homonyms will appear as different terms, connected by their identical spelling but differentiated by their definitions.

(d) The equivalent in other language does not mean — if possible — the translation of the term, but the term denoting the “X” language the concept expressed by the definition of the term as it occurs in “Y” language. If such a term does not exist, a translation is used identifying the fact that it is a translation and not an existing language equivalent.

As a consequence of the requirement, according to which the data base should contain further the equivalents of a term in the vocabulary of different schools of thought, each term is identified (if possible) by stating the context, such as school of thought, culture, etc. in the framework of which its meaning is defined. This is done, e.g. by using proper nouns as suggested by Mr. S. Klein (6).

(e) The data base contains references to the source

of the given term, i.e. to the document (terminological standard, vocabulary, encyclopaedia, textbook, article, etc.) which contains the definition, explanation and/or context of the term as it is included into the data base. The cases, when the definition or explanation included deviates from that of the source, as well as the cases of 'non-documentary' sources (e.g. definitions elaborated by an inputting organization or by a scientist for the reasons of Interconcept) are identified accordingly.

The required multiple use of the Interconcept data base can be assured if:

- services of both retrospective and current types (i.e. information retrieval as well as SDI on new concepts) are provided;
- parts as well as the whole of the machine readable file are output in order to serve different types of user needs;
- output is produced:
 - (1) via displays to serve single users;
 - (2) in printed form in order to reach a wide audience;
 - (3) on magnetic tape to be transmitted to other systems, which want to use it (eventually together with the appropriate instructions and software).

A few forms of output of special interest to indexers could be the following:

- retrieval of a single term in order to learn its definition and/or explanation, its relations to other terms, its correct translation to another official language of Unesco;
- retrieval of groups of terms according to a classification scheme;
- retrieval of interrelated terms by means of chained search;
- dissemination of information on new terms/concepts to interested institutions and scientists;
- publication of a new type of multilingual dictionaries containing not only literal translations of terms, but also their equivalents and/or approximates in different schools of thought.

The most important would be, however, to use the information available in the data base to construct analytic and systematic glossaries.

During the ISSC/COCTA workshop on the guidelines for conceptual analysis, it was pointed out that the analytical exercises involve two aspects or components, as follows (4):

(a) Reconstruction: the effort here is to identify concepts that have been defined "by scholars who use the selected term in a systematic way. Their definitions have been subjected to scrutiny in order to identify the main 'characteristics' used to establish the connotations of these concepts, and then to cluster them in such a way as to establish the conceptual distinctions that have been made".

(b) Construction: on the basis of the information generated by the reconstructions, an analytic glossary is compiled which identifies and gives terms to the various important concepts for which each of the key-terms has been used. The context of use of each of these concepts is also examined in relation to different disciplines and subject fields. Thus an output of each of the concept construction projects is expected to be an analytic glossary of limited scope.

The systematic glossary, on the other hand, is the

output of an effort to identify a group of 'associated concepts' which are central to the research and publications of a subject field, so that they are presented in a systematic way, with suitable definitions and the terms used by members of a discourse community to refer to these concepts can be identified.

4. Organizing the system

As a consequence of the fact that Interconcept is supposed to provide information both on currently used ("old") and newly introduced concepts, the creation and updating of the data base requires partly retrospective, partly continuous collection of data for input purposes.

It should be recognized further that input in the Interconcept system and input in the computerized data base are two different things. In some cases the information received or extracted from outside sources will be complete enough to be input right away in the data base (a term together with its definitions and relations to other terms). More often, however, the first data received from outside sources can be considered only as "raw materials" of the system.

The sources of the retrospective collection of data could be the following (in order of priority):

- research works on a specialized field;
- terminological standards;
- files of terminological data systems;
- terminological vocabularies, dictionaries, encyclopaedias;
- textbooks and manuals;
- thesauri, classification systems, etc.

The sources of the continuous collection of data may be:

- the sources already mentioned in connection with the retrospective data collection, published after establishing the retrospective file;
- primary social science literature (e.g. research reports, studies, articles; proceedings, discussions and recommendations of congresses, conferences, symposia) — and also the documentation services which process it;
- contributions by institutions, scholars and other specialists in the social sciences.

The manifold and sophisticated structure of the subject fields of Interconcept as well as the special knowledge required, the amount of work to be done, and the widely spread sources the system has to use, all postulate the division of labour necessary to provide the input to the system. On the other hand, the international aim of Interconcept, namely to develop international understanding and cooperation in the field of social sciences, may be achieved only if the output of the system is used internationally.

It is characteristic for the development of information systems in the last years that the sphere of users becomes more and more identical with those providing the systems' input: the systems are established on the basis of division of labour; the points of the system are providing the input for the central processing unit with the aim of obtaining the system's entire output. The international information system for Interconcept will be established according to the above principle.

5. Some findings of the Pilot Project

In the framework of the Interconcept Pilot Project, Mr. F. R. Riggs undertook the analysis of a sample of abstracts

prepared for the World Congress of Sociology held in Uppsala, Sweden, in August 1978. The purpose of this exercise was to see whether an analytic glossary could be "... employed to sharpen the focus of an indexing system and also to demonstrate the prevalence of ambiguity in the way important terms are used by social scientists".

(2)

One of the aims was to find out how many concepts of 'development' were used in each abstract. Mr. Riggs found that 17 out of the 40 abstracts used only 1 concept, and 15 abstracts used 2 concepts of 'development'; 6 abstracts even used 3, and 4 abstracts used as many as 4 concepts.

It was interesting to find out how the traditional indexing of these abstracts compared with an indexing system built on the concepts studied rather than on the terms used. The question posed was whether the index term 'social development' matches the concept of social development as defined for the purpose of our Pilot Project. On investigation, it was found that only 9 out of 20 abstracts dealing with the concept of social development were also indexed under the term 'social development'. On the other hand, 9 abstracts dealt with the concept but were not indexed under 'social development'. Four of those indexed under 'social development' did not deal with this concept.

The above findings form naturally only a small part of the exercise undertaken in the framework of the Interconcept Pilot Project and the improvement of indexing is only one of the aims of Interconcept. Nevertheless, we feel that even our small scale project and the experiments undertaken on the basis of a small sample give evidence of the impact Interconcept might have on the development of indexing in the field of social science literature.

References

- (1) Interconcept: the establishment of an international information system of social science concepts. Background paper to the Meeting of Experts on 'Interconcept' principles and strategies. Paris, 1977.
- (2) Riggs, F. W.: Final report to Unesco on the Interconcept Pilot Project. Honolulu, Hawaii, 1978.
- (3) Riggs, F. W.: On the meanings of 'Development' - A summary statement for the Unesco Interconcept Pilot Project. September 1978.
- (4) Riggs, F. W.: Report on Skokloster Workshop. September 1978.
- (5) Doyle, B.: Information Retrieval and Processing. Melville Publishing Co., Los Angeles 1975.
- (6) Klein, M. S.: Toward terminating the terminological turmoil: the proper nouns solution (Intern. Studies Ass. Working Paper No. 67, April 1976. Univ. Center for Intern. Studies, Pittsburgh).

SPECIAL NOTICE

BSO Switching Exercise

FID and UNESCO have under consideration the possibility of carrying out a test-exercise to secure factual information upon the ability of the "Broad System of Ordering" to serve as a switching medium, enabling document entries indexed in one indexing language to be made available indexed in a different indexing language.

Enquiries are invited from special libraries or documentation centres which might be interested in taking part in such an exercise, which will consist of a two-way exchange of indexed document entries between two centres. Each centre would supply to the other, via the FID/BSO Panel (which will undertake index switching) a sample of document entries indexed in its 'home' indexing language. Subsequently each participating centre would receive, indexed in its 'home' indexing language, the sample of document entries originally indexed by the other centre in the latter's indexing language. Each centre would be asked to supply a report on the material it receives, covering such questions as the extent to which it is helpful to have the switched information (a) in any form (b) in the form as actually converted by the switching process, and the effort, nature, and likely cost of further processing at the receiving centre which might improve its usefulness.

It is envisaged that the two centres participating would not have identical focal-interest (or core) subjects, but that they would have appreciable areas of marginal or fringe subjects in common.

The sample contemplated would be the product of the indexing of 500 documents of current input at the centre concerned.

It is emphasised that in this experiment, the roles of the centres would be confined to supplying copy of their indexed input and to preparing an evaluative report upon the switched material received. Centres would not be called upon to classify document entries by BSO. The actual switching would be undertaken by the FID/BSO Panel.

Further details, and preliminary questionnaire form may be obtained from FID, P. O. Box 30115, 2500 GC The Hague, Netherlands.

(From: FID News Bull. 30 (1980) No. 1)

Mr. Einar Öhman, 1896–1980

It is with deep regret that we inform you of the death of Mr. Einar Öhman on Saturday 12 January 1980 at his home in Avägen 4, Täby, Sweden.

Mr. Öhman was long interested in classification matters and in particular the Universal Decimal Classification (UDC) and at an early date was involved in FID/UDC affairs.

In 1958 he was elected a Vice-President of FID for the period of 1959 through 1962. He resigned as a member of the FID Council in late 1961 when he was appointed Head of the FID Classification Department. Mr. Öhman remained in this post until his second retirement in December 1966 at the age of 70 years.

Subsequent to this Mr. Öhman remained very active in the affairs of the FID/CCC "Central Classification Committee" for the development of the UDC of which he was a honorary member.

He was the author of several articles on classification amongst them being "Alphabetical Indexes to Chinese Classification Systems" presented at the 1976 FID Symposium held in Commemoration of the Dewey Centenary.

General Secretariat of the FID