

SEMINAR ON THESAURUS (1975). Paper BD

WORK AND TIME ANALYSIS FOR RENDERING TERMS IN A THESAURUS DERIVED FROM A FACETED CLASSIFICATION SCHEDULE -- A CASE STUDY WITH LATHE PRODUCTION

K N PRASAD, S K MAHAJAN, and J THYAGARAJAN,
Central Machine Tool Institute, Tumkur Road, Bangalore 560 022

Describes the various steps involved in rendering a standard term. Identifies the difference in the role of the term occurring in a faceted classification schedule. Analyses the work involved in the conversion, and the time required for each of the steps.

0 INTRODUCTION

01 Thesaurus

A thesaurus is a controlled but dynamic vocabulary of semantically related terms, offering comprehensive coverage of a domain of knowledge. Its main use is in the subject characterization of documents and queries in information storage and retrieval systems based on concept coordination. Its principal elements are descriptors, non-descriptor terms, and relationship indicators. It generally comprises one or more systematic displays and one or more alphabetical lists. (4)

02 Thesaurus Types

There are different types of thesauri, ranging from purely alphabetical arrangements to varying combinations of alphabetical lists with systematic displays. They are : 1) Alphabetical thesaurus ; 2) Alphabetical thesaurus with broad subject classification; 3) Alphabetical thesaurus with clusters and arrowgraphs ; 4) Alphabetical thesaurus with hierarchical displays ; 5) Alphabetical thesaurus with hierarchical classification; 6) Alphabetical thesaurus with faceted classification-broad groups ; 7) Alphabetical thesaurus with faceted classification-detailed ; and 8) Systematic thesaurus with alphabetical index (1).

03 Faceted classification

A freely faceted classification is a scheme for classification guided by explicitly stated postulates, canons, and principles for the three planes of work - and involves the analysis of the subject into its facets in the idea plane, their transformation into kernel terms in the verbal plane using current standard terminology, their translation into kernel numbers in the notational

plane according to a scheme of classification, and the synthesis of the kernel numbers into a class number. Such a scheme for classification with equal facility, co-extensively classify a microsubject, such as that usually embodied in articles in periodicals, technical reports and patents, as well as a macrosubject, such as that usually embodied in a whole book. This methodology, being based on postulates about the structure and development of the universe of subjects, theoretically admits of the addition of a large number of characteristics one after another in a large number of facets to be added one after another in the classification of the universe of subjects (3).

04 Development Plan

It is known experience among indexers, that each technical information centre may have to design its own thesaurus for use in indexing/ searching of a file of information. Therefore, several attempts have been made to standardise the procedures for thesaurus construction for information retrieval purposes. This paper gives a rough estimation of time taken in preparing a worksheet for computer generation of thesaurus, with faceted schedule on lathe as a basis.

I SCOPE OF THE PAPER

The purpose of this paper is to describe the various steps involved in rendering a standard term. It identifies the difference in the role of the term occurring in a faceted classification schedule. The work involved in this conversion, and the time required for each of these steps is analysed.

II Boundary Conditions

The following are the main boundary con-

ditions within which the faceted thesaurus could be created.

- 1 Knowledge of the methodology of designing a scheme for classification;
- 2 Knowledge of principles and processes of the theory of library classification ; and
- 3 Knowledge of the principles and processes of thesaurus construction.

PROCEDURES ADOPTED FOR THE DESIGN

20 Steps

The main steps involved in the conversion of a faceted classification into a faceted thesaurus

- '1 Study of the existing schedule, regarding
- 11 Its comprehensiveness;
- 12 Arrangement of quasi-isolates and speciators within each of the quasi-isolates; and
- 13 Irrelevant quasi-isolates and spectators.
- 2 Recording of terms
- 21 Incorporating synonyms, near synonyms and alternate word forms;
- 22 Writing of scope notes and definitions (where necessary) and
- 23 Verification with respect to term representation, viz., noun form, singular/plural form, abbreviations etc.
- 3 Writing devices wherever used
- 4 Assigning Level numbers
- 5 Assigning Class numbers
- 6 Final verification and editing
- 21 Study of the existing schedule

The faceted classification schedule for lathe production designed by Gopinath (2), was taken for conversion into a faceted thesaurus.

The schedule was carefully studied, especially its structure. It was also found that the schedule was not comprehensive enough to cover the latest literature on lathes. Secondly, some of the quasi-isolates were found to be inappropriate. In some cases the speciators used were not relevant for the purpose. New quasi-isolates were introduced wherever found necessary.

22 Recording of Terms

The terms used in the classification schedule were taken, incorporating the changes wherever found necessary. The steps involved in it were :

- 1 Quasi-isolates were rearranged;
- 2 The speciators were written with their synonyms wherever available, separated by a 'slash' (/); and
- 3 The UNISIST's guidelines were followed for the presentation of terms in the thesaurus (6).

221 Word Form

The term 'By' which precedes each of the quasi-isolates in the schedule, was dropped in the thesaurus.

Inverted form of the terms were used wherever found necessary.

Example : 'Bearing design' instead of 'Design of Bearing' which is the natural order of presentation. Cross reference is given from the lead-in-entry.

Inverted entries are not possible in many cases.

Example : 'Number of spindles' when inverted gives 'Spindles Number' which does not convey the meaning, it was intended to convey. In such cases the natural order had to be maintained.

In the classification schedule, 'By power' is a quasi-isolate and 'Manual', 'Mechanical', etc., are the speciators under it. But the terms 'Manual', 'Mechanical' themselves do not convey any meaning in the thesaurus, when they are individually highlighted. In order to make the sense complete, in the above example, 'Manual', 'Mechanical¹' etc. are re-written as 'Manual Power', 'Mechanical Power' in the thesaurus.

222 Synonyms, Near Synonyms and Alternate Word Forms

The synonyms collected were added to the preferred terms with a slash in between. In some cases there were more than one synonym for a preferred term. These were indicated as shown in the example.

Example : * Safety clutch/slipping clutch
* Safety clutch/overload clutch

Near synonyms were included in some cases as lead-in terms.

Example : — Ultra precision lead screw for
High precision leadscrew

Wear compensation is taken as synonymous with Backlash compensation, though they are

not synonymous in the strict sense, because the elements used to bring about the desired effect being the same, and also backlash, being due to inaccuracies in the newly fitted leadscrew and the wear being introduced after a period of working. The device used to eliminate the first said defect is also used to eliminate the wear in lead-screw while in operation.

223 Acronyms

Acronyms in some cases were used as preferred terms since they are in general use and the potential users are familiar with them. Even in the documents they are used in the same way,

Example : PCD = Pitch Circle Diameter
 MT = Morse Taper

224 Spelling

Only the most widely accepted spelling of the word was adopted.

225 Noun Form

Terms have been written in noun forms but adjectival forms were also included, wherever found necessary.

Example: Mechanical, Axial

226 Number

Singular form was mostly used except where the proposed term was a count noun.

Example: Number of Bearings

23 Devices

The following devices have been used in the thesaurus:

- 1 Alphabetical Device (AD) - generally used for the Brand names
- 2 Chronological Device (CD) - used to represent the year of make
- 3 Geographical Device (GD) - For terms representing the country of make
- 4 Numerical Device (ND) - This device is used in many places. It represents the exact measure of a given attribute.
- 5 Subject Device (SD) - Used to get the speciators relating to materials.

24 Assigning Level Number

Level numbers were given according to the prescriptions laid by Messrs. Carolyn Watters and Michael Shepherd (5) except that the basic subject is given the level number 00.

25 Assigning Class Number

Class number in alphanumeric were assigned in addition to the level number.

26 Final Verification and Editing

- 1 Checking for correct representation of levels and notation;
- 2 Checking for word form, spelling etc;
- 3 To check, if the devices were properly used, and
- 4 To check if the slash was used to separate the synonyms

3 CASE STUDY

The following table illustrates the work and time analysis for the conversion of the depth schedule on lathe production into a thesaurus.

Sl. No	Nature of work	Quantity of work	No of personnel	Total Time taken
1	Study of depth schedule			3 days (30 hrs)
2	Discussion at DRTC			1 day (6 hrs)
3	Literature survey	5 Hand books 25 Other books 20 Reports 100 Trade catalogues 15 Periodicals		2 days (12 hrs)

SI. No	Nature of work	Quantity of work	No of personnel	Total Time taken
	Scanning and selection of terms	5 Hand books 10 Other books 10 Reports 100 Trade catalogues 100 Articles		20 days (175 hrs)
	Discussion at DRTC			1 day (6 hrs)
	Incorporation of new terms and synonyms ; Deletion of irrelevant OI's and speciators ; incorporation of new QI's and speciators			2 days (30 hrs)
	Discussion at DRTC			1 day (6 hrs)
	Re-writing of the schedule with the new terms	About 1 300 terms		5 days (50 hrs)
	Discussion at DRTC			1 day (6 hrs)
10	Assigning of levels and class numbers			5 days (50 hrs)
11	Typing of schedule			4 days (20 hrs)
12	Discussion at DRTC			1 day (4 hrs)
13	Incorporation of corrections subsequent to discussion			4 days (32 hrs)
14	Retyping			4 days (20 hrs)
15	Discussion at DRTC			1 day (6 hrs)
16	Final verification and incorporation of minor corrections			2 days (8 hrs)

4 ANNOTATION

1 At various stages of the conversion of the faceted schedule into a thesaurus, the help of DRTC was taken, and the useful discussion with its faculty reduced our workload to a considerable extent.

2 The team which comprised of two subject specialists and one documentalist, needed very little time for selection of terms and their subsequent arrangement. It was also easier, for deleting irrelevant quasi-isolates and speciators and incorporating new quasi-isolates and speciators

3 The knowledge of the theory and practice of library classification and the knowledge of the principles and processes of thesaurus compilation helped in the conversion process.

4 At onestage it required to make some major corrections which consumed 32 man-hours.

5 Since the structure of nascent subject is frequently liable to change on account of developments in the wavefront of that subject, the thesaurus has to be made up-to-date. This is a continuous process and the user's interaction over the same has to be kept in view. This process cannot be quantified, and therefore, the time factor cannot be determined.

5 CONCLUSION

The conversion of the lathe schedule into a thesaurus was taken up as a pilot project. The total time taken has been 52 days only, since the lathe schedule designed by Gopinath (2), was a fairly comprehensive one although it needed to be updated. Based on the completion of this project, we are finally compiling a thesaurus for the complete field of machine tools. The time that is required for completion of this whole project will be considerably less for two reasons :

1 Many of the terms that are incorporated in the lathe schedule are common to other machine tools ; and

1 Possession of knowledge of compiling such a thesaurus,

On the other hand there is also the possibility of requiring more time, as the other machine tools (excepting Grinding machines and Milling machines) do not have depth schedules for classifications.

In the field of Machine tools, most of

the terms have synonyms and the selection of these poses a big problem. It requires more time. What is more essential is to keep the thesaurus up-to-date which necessitates continuous scanning of relevant literature.

6 ACKNOWLEDGEMENT

We thank the DRTC faculty for their kind cooperation and continuing guidance. We also thank our Director for permitting us to present the paper to this seminar.

7 BIBLIOGRAPHICAL REFERENCES

- 1 Sec 02 AITCHISON (Joan) and GILCHRIST (Alan). Thesaurus construction: A practical manual. 1972 P. 50-66.
- 2 Sec 21, 5 GOPINATH (M A). Lathe production: Depth Classification Version of CC (Lib Sc 8; 1971 ; K)
- 3 Sec 03 NEELAMEGHAN (A). Theory of classification. {In KENT (Allen) and LANCOUR (Harold), Ed Encyclopaedia of library and information services. Vol 5 1971. P 160)
- 4 Sec 01 NUCLEAR INFORMATION (International -- System). Manual for indexing 1970. P 9)
- 5 Sec 24 SHEPHERD (Michael) and WATTERS (Carolyn). Computer generation of thesaurus. (Lib sc. 12; 1975; E)
- 6 Sec 22 UNINIST. Guidelines for the establishment of monolingual thesauri. 1973.