FIELD DOCUMENTATION:
A TECHNICAL REPORT ON INSDOC - NML PROJECT

S PARTHASARATHY
S DUTTA
A S RAIZADA
M RAMACHANDRAN

Insdoc, Delhi-12.

1 INTRODUCTION

For the first time Insdoc launched on what is known as Field documentation. It consisted in conducting a new experiment by setting up a Central Information Index using Universal Decimal Classification. For this purpose a team was sent to National Metallurgical Laboratory, Jamshedpur. The organisational details have been described elsewhere in a separate report. This attempt to present the technical aspect of the work involved in the experiment.

1.2 In situ Documentation

Documentation service rendered at the point of actual consumption may be termed as in situ Documentation. The Central Information Index is an example of in situ Documentation. Such a service functions in close cooperation with the scientists and research workers. It is tailored to their immediate needs. It is modified by their feedback.

1.3 Scope of the Project

1 to rationalise the practices followed by the scientists working in the Laboratory for compiling literature references for their use;
2 to pool their efforts in the selection and preparation of abstracts of articles relevant to their projects;
3 to compile a Central Information Index comprising of the references selected and abstracts prepared by the scientists relating to their projects;
4 to classify the entries minutely according to UDC for arranging them; and
5 to prepare an analytical subject index, using chain indexing technique for preparing subject entries.

2 SUBJECT

The documents in metallurgy - mostly articles appearing in scientific and technical periodicals - are multifocal, and multifaceted. The following table represents the occurrence of facets in a preliminary subject analysis of a random sample of 300 articles in the field of metallurgy:

<table>
<thead>
<tr>
<th>No. of Facets</th>
<th>No. of articles</th>
<th>% of articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>1.3</td>
</tr>
<tr>
<td>3</td>
<td>32</td>
<td>10.6</td>
</tr>
<tr>
<td>4</td>
<td>124</td>
<td>41.3</td>
</tr>
<tr>
<td>5</td>
<td>90</td>
<td>30.0</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>13.3</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>1.6</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>1.3</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>Total</td>
<td>300</td>
<td></td>
</tr>
</tbody>
</table>

Thus, from the above table it is clear that majority of the articles present four facets. 71% of the articles presented four to five facets (Table 1). It can be safely inferred that a classifier has to be prepared to accommodate four to five facets in the class numbers assigned by him.

3 CLASSIFYING

It was found helpful to analyse articles and synthesise this result of facet analysis to get the Class Number. This method gave enormous consistency over a range of 2,500 assorted articles.

Our experience shows that a trained classifier can classify and check six articles per hour, on the basis of the turnover of two classifiers for eleven days. The initial take
off period is three days. In this period familiarity with the schedules and with the ramifications of the subject metallurgy is acquired. The average curve smoothens and appears to be stabilised with the average production of 40 entries.

4 CLASSIFICATORY PROBLEMS

For efficient indexing it becomes necessary to assign coextensive class numbers to the documents or articles. This becomes all the more necessary as the primary purpose of classification at micro stage is to throw the entries in an helpful sequence – a sequence from which it becomes possible to retrieve a group, howsoever deep, with minimum of noise factor.

4.1 Coextensive Class Numbers

All articles were minutely classified giving co-extensive class numbers. Most of the articles treated a combination of subjects demanding multifaceted numbers. In UDC we have to use large blocks of numbers to represent the subjects. UDC does not prescribe any fixed sequence for these blocks or facets. It provides for alternative placings for each of these facets. This implies at least four entries for a four-faceted subject, five for a five faceted subject and so on. Most of the articles in metallurgy consisted of four to five facets. For 2,500 articles, we would require more than 10,000 entries. With the increase in the number of articles, the size of the Central Information Index would increase enormously and its size alone would hinder its usefulness. In order to avoid this shortcoming, it was decided to give only one entry for each article, instead of many entries for the various aspects. To make this effective, it was necessary to fix the sequence of the facets in a class number and provide for all the alternate approaches through analytical subject headings.

It was not possible to assign coextensive class number to each and every document. The following table shows the results of the attempts to arrive at coextensive class numbers in a representative sample of 300 articles.

<table>
<thead>
<tr>
<th>No. of Facets</th>
<th>No. of articles</th>
<th>No. of articles having coextensive Cl Numbers</th>
<th>Percentage of articles to which coextensive Cl No. could be assigned</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>3</td>
<td>75</td>
</tr>
<tr>
<td>3</td>
<td>32</td>
<td>18</td>
<td>56.2</td>
</tr>
<tr>
<td>4</td>
<td>124</td>
<td>51</td>
<td>39.8</td>
</tr>
<tr>
<td>5</td>
<td>90</td>
<td>27</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>11</td>
<td>27.5</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>300</td>
<td>110</td>
<td>36.6</td>
</tr>
</tbody>
</table>

Table 2 represents the distribution of the articles for which coextensive class numbers have been assigned. It will be seen that with UDC, it has been possible to assign coextensive class number fairly satisfactorily to those articles presenting upto four facets.

4.2 Limitations

The limitation in representing all the facets in a class number i.e. in designing coextensive class numbers is traceable to two causes. First cause is that a classifier has to grapple with the wave front of knowledge where everything is transient. New concepts appear as if they are not settled in relation to the subject. Hence although it is possible to spot a new facet in an article, it becomes very difficult to ascertain its proper relationship with other facets. Second cause is due to shortcomings of the schemes. UDC failed on two counts. One was that many of the isolates were not enumerated. It is not possible for any scheme to enumerate future isolates in any subject. The second reason was that with the increase in the number of facets the corresponding coextensive class number becomes too long - so long that it distracts the attention of the user.

4.3 Compromise

The efficiency in the Central Information Index was obtained with the aid of compromise between co-extensiveness on the one hand and length of class number on the other. Verbal extension was freely used to sharpen the isolates. One such case was 'Impact
extrusion'. The enumerated isolate in the schedule was 621.77 Extrusion, rolling etc. The verbal extension method gave the isolate number for Impact Extrusion as 621.77 [1]. Thus the class number was made coextensive without increasing the length proportionately. The other method was to strike a balance between the class numbers and subject-index. By this, the number of facets were deliberately kept under six and the rest of the facets were delegated to the subject index. It must be pointed out that these practices mark a definite departure from UDC's conventions.

Just to give an idea of the length of class number an example is quoted below:

622.765:622.775:622.343.5:66.094.3:669.537.015.9

Author KLEBONOV (O B) et al.
Title Use of "clinker" of electrolytic zinc plants for precipitation of copper in beneficiation of oxidised copper ores by the combined system.
Citation Tsvetn. Metal. 35, No. 11, 1962, 36-37.
(Chem Abs. 58(13)24 June 1963, 13495 a)

Abstract
Zn plant "clinker", contg. Fe total 32.5 and Fe metallic 31.6%, and its magnetic concentrate replaced satisfactorily Fe sponge and filings as a Cu precipitant in the treatment of oxidized Cu ores by the leaching-cementation-flotation method. Overall Cu extn. into the intermediate product increased to 78, and into finished Cu concentrate to 69%. There were many cases that called for the use of as many as three analytical subdivisions of 669. One is forced to use colon and thereby increase the length of class numbers. Whereas had there been a provision for using these analytic subdivisions without repeating 669 and using a colon the class number would have been reduced to:

resulting in an economy of six digits. Here 0 and hyphen serve only as connecting symbols. This definitely is as coextensive a class number as the earlier one, with lesser number of digits. There were many such cases.

4.41 Long Class Number

Many articles called for long class numbers. One such case is cited in sec 4.3 as an illustration. Many factors add to the length of the class numbers. The use of relation sign i.e. colon as a means of sharp-ening an existing isolate is one such factor. One has to resort to this method owing to lack of enumeration.

There were many cases that called for the isolate numbers both from 669-1/-9 and 669.01/.09. Since there appears to be no provision of using these isolate numbers in one and the same class number consecutively, the use of colon is inevitable. Thus the class number for

"Conditions of dissolution and removal of oxygen during deoxidation of metal Armco Fe melts".


This article calls for the use of as many as three analytical subdivisions of 669. One is forced to use colon and thereby increase the length of class numbers. Whereas had there been a provision for using these analytic subdivisions without repeating 669 and using a colon the class number would have been reduced to:

resulting in an economy of six digits. Here 0 and hyphen serve only as connecting symbols. This definitely is as coextensive a class number as the earlier one, with lesser number of digits. There were many such cases.

4.43 General Vagueness

A case was that of aluminising of steel and iron. There are alternatives provided in the schedule. These are as under

4.4 Problems in UDC

The printed schedules of UDC (BS 1000: 1949, 669 Metallurgy) presented many problems. These problems manifest themselves acutely when an attempt is made to assign a co-extensive class number to a document.

4.41 Long Class Number

Many articles called for long class numbers. One such case is cited in sec 4.3 as an illustration. Many factors add to the length of the class numbers. The use of relation sign i.e. colon as a means of sharp-
669.198 Coatings on, or with Iron & Steel.
669.718.67 Coatings of other metals with aluminium.

4.44 Sharpening of existing subdivisions

In the absence of any guidance for sharpening any existing isolate number, the only alternative left to classifier is to use colon and then a chunk of numbers. This method is well illustrated in the schedules. This method presupposes that a new concept calling for a new isolate number is not actually a new concept but in fact a hybrid concept whose components are already enumerated somewhere else in the schedules. No synthetic number could be got for "Impact Extrusion" even by this method. The nearest number enumerated was 621.77. There were two alternatives. One was to classify all articles on Impact extrusion, and we had compiled a bibliography of nearly 120 articles, under 621.77. By doing this in a Central Information Index, which cumulates continuously we would have jeopardised the chances of subsequent retrieval of articles on Impact extrusion. At least there would have been a considerable noise factor. The other was to ignore the general directions of UDC and find some device to extend 621.77 in a way that it does not come in conflict with future expansion of UDC. The device used was that of "verbal extension", e.g. 621.77 [I) for Impact extrusion. But this device too has a limited use. It was failing even in this case. Articles on cold impact extrusion had to be separated from other articles on Impact extrusion in order to minimise the noise factor. The only alternative left was to use alphabetical device or to subdivide 621.77 [I]. The former was preferred. Another case was when a number had to be found out for "slugs". The nearest enumerated number was 669-412 Ingots, slabs etc. 669-412 [1] slugs was used. The absence of general principles for sharpening the existing subdivisions was felt throughout our work.

4.45 Two Uses of Colon

It was found at the stage of arrangement of entries that colon has been used in two ways. One as a means of sharpening an enumerated subdivision. The other use is that of showing relations between two class numbers. The filing position of colon has been defined. It is in the former use that the arrangement becomes from specific to general instead of from general to specific. In this case the enumerated isolate numbers appeared later than that of synthetic isolate numbers. It became impossible to rectify this without separating the duality of colon. And it was not attempted.

4.5 UDC Practice

UDC appears to advise against the construction of long numbers. It may be remarked that the length of the number cannot be shortened by instructions alone, as the length of any class number is not a function of a scheme only. It is more closely related to the subject of a document. Another suggestion is to use relation sign (:) colon between two or more facets, to make as many copies of an entry as there are facets and file one copy under each facet. This is an ideal solution when the number of facets presented are two. For metallurgical literature, as stated earlier, majority of articles present 4 facets, and keeping four copies of every entry makes the index unwieldy. In fact for incorporation of 2,500 entries in the Index, the team would have had to prepare 10,000 entries in the classified part. Considering the labour involved, it was decided to have one entry per article in the main classified part of the index. The other approaches were to be satisfied by the subject index. This was another point of departure from the set practice of UDC which yielded an economy of 7,500 cards.

4.6 Unsolved Problems

It was difficult to reflect the relationships in the class numbers. The normal relations sought were utility, raw materials, property, process and equipment. The chunks represented these facets but did not indicate their relationship. In other words, the class number did not say whether 669.71 is a semi-finished or raw aluminium. However, the following principle gave enormous consistency:

Utility, Raw Material, Property, Process, Equipment

A similar problem came up in the field of chemical estimation. Here too, we put
chunks of numbers representing the substance analysed; the type of analysis and the element estimated. These facets representing the element or metal that has been estimated, the facet representing the method of estimation and the facet representing the sample. It became difficult to ascertain the sequence of facets from discussion. An 'ad hoc' sequence was followed and is given below:

Sample, method of estimation, element estimated.

Even in this, it was the 'sample' facet that gave considerable difficulties. One difficulty that could not be solved was "estimation of certain element in presence of other elements in a sample".

5 FEATURING

The classified part has to be duly featured. Co-extensive class number was given as far as possible for each entry of an article. All the entries were arranged strictly according to these class numbers. For feature headings, instead of using the co-extensive class numbers in full, only broad numbers followed by appropriate verbal extensions were used. All the verbal extensions following the same class number were arranged according to their subject filiation as represented by their full class numbers and not according to the alphabetical sequence of the verbal extensions. For analytical subject entries, for ease of location, only broad class numbers followed by verbal extensions were given. The above procedure, while securing the advantages of a completely classified arrangement of entries minutely classified, enables easy reference by using short class numbers followed wherever necessary by verbal extension. The purpose of class numbers is to mechanise arrangement as well as to enable expeditious retrieval. Minute classification naturally tends to give long class numbers which though helpful for arrangement are not helpful for retrieval. An attempt has been made in this project to secure efficient retrieval by using short class numbers with verbal extensions as location numbers for subject headings. This approach has provided good results in two of our demonstrations. A deviation from the normal method of preparing a feature card was felt necessary. This deviation was on account of the explicit realisation that storage mechanism is more or less independent of retrieval me-

chanism, and all the intricacies of storage mechanism need not be reflected in the retrieval mechanism. Hence it was decided to have only 10 to 12 digit class numbers on the feature headings. These feature heading cards were termed as main feature cards. Deeper feature headings carried no class numbers but only the terms. Their position was of course determined by the class numbers. These were called as sub-feature headings or merely sub-headings. Each main feature card gave the arrangement of sub-headings. Thus a main feature card looked like:

<table>
<thead>
<tr>
<th>Class Numbers</th>
<th>STEEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub-headings</td>
<td></td>
</tr>
<tr>
<td>Mechanical property</td>
<td></td>
</tr>
<tr>
<td>Embrittlement</td>
<td></td>
</tr>
<tr>
<td>Solution</td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td></td>
</tr>
<tr>
<td>Testing</td>
<td></td>
</tr>
</tbody>
</table>

Class numbers on the feature cards serve no purpose other than their replacement in the classified part. A reader, who has changing interests, is allergic to long class numbers.

The method of featuring followed is a synthesis of verbal extension and chain procedure. It has three advantages:

1. It does not load each feature heading with long and complicated class numbers;
2. It makes the location in the classified part comparatively easy; and
3. Readers' attention is less distracted.

6 SUBJECT INDEX

The alphabetical subject index to the Central Information Index was made on 75 x 125 mm cards. As far as possible the entries for the subject index were derived by chain procedure. On account of majority of non-coextensive class numbers, it became necessary to scan the abstracts to the articles and select ideas not covered in the class numbers.

A deviation had to be made in respect of index numbers. Normally the index number should be the class number. But as it became so unmanageably long that it lost its value as index number. Therefore, the index number did not indicate any specific article but it
referred to the group containing the article. The directing element had to be changed from 'see' to "See under".

6.1 Illustration

The following examples show the mechanism and the departure from the set practices of chain procedure for preparing a subject index:

<table>
<thead>
<tr>
<th>Terms used</th>
<th>Actual number</th>
<th>Index Number used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cast iron</td>
<td>669.13</td>
<td>669.13</td>
</tr>
<tr>
<td>Grey cast iron</td>
<td>669.131.6</td>
<td>669.131.6</td>
</tr>
<tr>
<td>Inclusion, Grey cast iron</td>
<td>669.131.6:620.192</td>
<td>669.131.6:620.192:669.046:58</td>
</tr>
</tbody>
</table>

7 ASSESSMENT OF WORK LOAD

The work relating to the Central Information Index falls into the following categories:

1. Selecting relevant references and abstracts relating to project;
2. Preparation of abstracts wherever necessary;
3. Typing of references and abstracts on standard 125 x 200 mm slips;
4. Comparison of typed slips;
5. Classification of the entries;
6. Preparation of subject index entries and compilation of authority list;
7. Classified arrangement of the entries and preparation of the feature headings;
8. Alphabetic arrangement of subject entries; and
9. Maintenance of Central Information Index.

The work relating to the first two items is done by scientists working on the various projects. The work pertaining to items 4 to 9 is to be done by trained documentalist. The typist does the work pertaining to item 3. He has to be trained for this job. On an average a typist can type 10 abstracts or 20 references per hour. The work pertaining to items 5 and 6 falls within the purview of documentalist. Even though there is bound to be some increase in the output as the documentalist gains experience and gets familiarity with the subject and classification procedure, a steady state in the output is reached fairly early. Maintenance jobs entail a routine of checking up the classified sequence, the alphabetical sequence and replacing worn out entries. This has to be done daily.

8 CONCLUSIONS

The majority of the articles in metallurgy called for the use of five facets only. Out of these a four-faceted class number could accommodate the maximum number of articles. This experiment points to the necessity of studying the problem of indexing in its entirety. The experiment gave a valuable experience. The time for conducting similar experiments in other fields of knowledge has arrived.

9 ACKNOWLEDGEMENT

We are grateful to the Director-General, Scientific and Industrial Research, for his encouragement and support for initiating this project, to the Director, the Deputy Director and the Scientists of the National Metallurgical Laboratory for providing us all facilities and giving us encouragement and technical guidance in the organisation of the Central Information Index.