CHAPTER 8

SUPPLEMENTARY INDEXING

At the same time as we were arranging for the compilation of questions, we invited those who were co-operating in this task to index documents by any of the systems being used in the project. The purpose of doing this was in order to have some data which would provide a comparison with the indexing done by the project staff. Those who agreed to help (see Appendix G of Ref. 1) were sent selected lists of documents which had been indexed in the project, and relevant indexes to the particular system or systems which they were using. Those indexing by U.D.C. received the alphabetical index to U.D.C. numbers which had been compiled by the project staff of all terms used in the indexing of the first 10,000 documents, together with information concerning the printed schedules which were being used. A list of the alphabetical subject headings or a list of the uniterms used in indexing the first 10,000 documents was sent to those indexing by Alphabetical or Uniterm, and a set of the schedules, the alphabetical index and indexing rules were sent to those using Facet. Each person also received a set of master indexing cards. (see Apprendix 8A).

A total of 3,793 items were indexed, and Table 8.1 gives relevant data on this indexing.

	No. of documents indexed	Average Indexing Time	Average entries
U.D.C.	1,115	10.8	3.6
ALPHABETICAL	902	11.3	1.1 (with 5 elements)
FACET	775	10.8	12
UNITERM	1,001	8.0	3.6

TABLE 8.1

DATA ON SUPPLEMENTARY INDEXING

Comparison with the indexing done by project staff showed that the average time fell near the middle time level used in the project, and that the average number of entries for each document was very close to that requested by project staff when indexing at 8 minutes. Further statistical data on the supplementary indexing is given in Chapter 7 of Ref. 1.

It was originally contemplated that the indexing entries would be incorporated in the project catalogues, but this was not done, partly because it would have caused extra complications but mainly because it was found that checking of the efficiency of the indexing could be done more simply.

Many of the documents which had been indexed could not be included in the test because no questions had been received for those documents, but altogether it was possible to test the indexing of 1772 documents. The method of testing was to take the master search cards which had been used in the main test. One person (A) would have this search card while a second person (B) had the appropriate master indexing card, examples of which are shown in Figure 3. 'A' called out the first search programme and 'B' checked the master indexing card to see if the programme fitted the indexing. If successful, that completed the operation; if unsuccessful, 'A' called at the second search programme for 'B' to check, and so on. In cases where, in the project the search had been successful at an early stage but the supplementary indexing search had not succeeded, then 'A' would, whenever possible, devise further search programmes, but without knowing the indexing decisions entered on the master card. The results of this testing are given in Table 8.2 where comparison is made with the results for the source documents achieved by the project indexing. The reason for the total number of documents not agreeing is because many documents were indexed by two or more organisations. From this table project staff would appear to be superior in both U.D.C. and Alphabetical, but not so successful by Facet or Uniterm.

Table 8.3 breaks down the supplementary indexing by countries, and the results cannot, in view of the comparatively small number involved, be taken to show any marked national differences of the kind which might be expected to appear because of the popularity of different methods in England and the United States. Table 8.4 continues this breakdown to cover different organisations.

	SUCCESS		FAILURE		% SUCCESS	
	Project	Supplementary	Project	Supplementary	Project	Supplementary
U.D.C.	263	430	51	135	83.8	76.1
ALPHABETICAL	306	389	51	121	85.7	76.3
FACET	142	248	50	65	73.9	79.2
UNITERM	225	337	55	47	80.4	87.8

TABLE 8.2

COMPARISON OF RESULTS OF PROJECT STAFF AND SUPPLEMENTARY INDEXERS

	United States	United Kingdom	Canada	Holland
U.D.C.				
Success	44	370	-	16
Failure	14	112	-	9
% Success	76%	77%	-	64%
ALPHABETICAL				
Success	165	156	57	11
Failure	52	57	5	7
% Success	76%	73%	92%	61%
FACET				
Success	46	190	-	12
Failure	8	51	-	6
% Success	85%	79%	-	67%
UNITERM				
Success	206	115	-	16
Failure	23	21	-	3
% Success	90%	85%	-	84%

TABLE 8.3

SUPPLEMENTARY INDEXING RESULTS BY COUNTRY

Key a. Indexing Time in Minutes
b. Number of Documents
c. % Success

	U.D.C.			ALPHA.			FACET			UNITERM		
	a	b	c	a	b	c	a	b	c	a	b	c
UNITED STATES												
1.	-	-	-	15	5	80%	-	-	-	-	-	-
2.	-	-	-	9	20	85%	-	-	-	12	29	83%
3.	-	-	-	-	-	-	-	-	-	105	11	100%
4.	-	-	-	17	9	67%	19	7	43%	9	69	91%
5.	9	12	92%	8	13	85%	9	7	100%	5	12	92%
6.	-	-	-	12	9	78%	-	-	-	11	6	100%
7.	-	-	-	7	112	69%	-	-	-	-	-	-
8.	-	-	-	-	-	-	8	5	80%	5	13	85%
9.	-	-	-	13	11	91%	-	-	-	-	-	-
10.	-	-	-	6	4	100%	-	-	-	5	5	80%
11.	-	-	-	6	32	88%	-	-	-	-	-	-
12.	13	46	72%	-	-	-	12	35	91%	7	51	90%
13.	-	-	-	6	2	50%	-	-	-	7	13	85%
14.	-	-	-	-	-	-	-	-	-	8	20	95%
UNITED KINGDOM												
15.	36	27	89%	-	-	-	-	-	-	8	24	88%
16.	11	14	86%	-	-	-	-	-	-	-	-	-
17.	-	-	-	-	-	-	8	5	60%	-	-	-
18.	-	-	-	-	-	-	26	7	86%	19	11	64%
19.	21	11	82%	15	14	71%	15	7	86%	12	13	100%
20.	-	-	-	-	-	-	-	-	-	8	14	100%
21.	3	6	50%	-	-	-	4	6	83%	-	-	-
22.	14	18	61%	-	-	-	-	-	-	-	-	-
23.	-	-	-	-	-	-	6	29	79%	-	-	-
24.	11	13	85%	-	-	-	-	-	-	-	-	-
25.	-	-	-	3	67	70%	-	-	-	-	-	-
26.	10	30	77%	13	33	85%	24	17	94%	15	28	75%
27.	-	-	-	-	-	-	13	21	86%	-	-	-
28.	10	10	90%	-	-	-	-	-	-	-	-	-
29.	7	10	90%	-	-	-	-	-	-	-	-	-
30.	8	10	90%	-	-	-	-	-	-	-	-	-
31.	7	61	84%	8	28	50%	14	9	80%	5	10	80%
32.	19	9	67%	-	-	-	-	-	-	-	-	-
33.	-	-	-	-	-	-	14	41	83%	-	-	-
34.	10	21	71%	9	23	83%	-	-	-	-	-	-
35.	1	28	57%	1	25	76%	2	22	68%	1	27	85%
36.	8	54	65%	-	-	-	7	38	76%	-	-	-
37.	9	11	82%	-	-	-	-	-	-	-	-	-
38.	-	-	-	-	-	-	14	14	79%	-	-	-
39.	-	-	-	-	-	-	8	16	63%	-	-	-
40.	10	10	70%	-	-	-	9	7	72%	-	-	-
41.	7	14	79%	-	-	-	-	-	-	-	-	-
42.	5	13	69%	-	-	-	-	-	-	-	-	-
43.	14	16	88%	12	16	81%	-	-	-	-	-	-
44.	10	73	80%	-	-	-	-	-	-	-	-	-
45.	11	6	83%	-	-	-	-	-	-	-	-	-
46.	29	10	90%	8	7	86%	10	2	100%	2	9	89%
CANADA												
47.	-	-	-	5	53	90%	-	-	-	-	-	-
48.	-	-	-	8	9	100%	-	-	-	-	-	-
HOLLAND												
49.	29	25	64%	16	18	61%	23	18	67%	14	19	84%

TABLE 8.4

SUPPLEMENTARY INDEXING RESULTS OF INDIVIDUAL ORGANISATIONS

An analysis was made of all those cases where there had been failures with the supplementary indexing but where the project indexing had been successful. Not all the grouping included in the analysis of failures in Table 5.1 are appropriate, but they have been followed in Table 8.5, with some additional headings that were necessary. Table 8.6 groups these into the three main categories of indexing, searching and system, and it is not unexpected to find that indexing is the main cause of the failures.

		UDC	ALPHA	FACET	UNITERM	TOTAL
INDEXING						
2a.	Insufficient indexing	40	48	16	14	118
b.	Overdetailed indexing	1	6	1	1	9
c.	Incorrect indexing	2	1	2	0	5
d.	Lack of permutation	10	0	0	0	10
h.	Lack of knowledge of indexing practice	14	12	1	0	27
SEARCHING						
3g.	Insufficient searching	10	12	2	1	25
SYSTEM						
4a.	No. of places for same subject in schedules	6	0	4	0	10
b.	Lack of places in schedules	0	0	4	0	4
f.	Bad choice of heading	0	10	0	0	10
g.	Synonyms	0	0	0	3	3
h.	Inability to combine headings	0	13	0	0	13
j.	Lack of subject grouping of headings	0	4	0	2	6

TABLE 8.5

ANALYSIS OF REASONS FOR FAILURES WITH SUPPLEMENTARY INDEXING

	UDC	ALPHA	FACET	UNITERM	TOTAL
Indexing	67	67	20	15	169
Searching	10	12	2	1	25
System	6	27	8	5	46

TABLE 8.6

REASONS FOR FAILURES IN SUPPLEMENTARY INDEXING

From Table 8.5 has been extraced information relating to the efficiency in relation to the time spent on indexing, with the results shown in Table 8.7. In that the indexing times used by the project staff are estimated to have been at least 50% less than the equivalent time in a real-life situation, the breakdown of times in this analysis has been so adjusted to correspond as far as practical to the indexing times used in the project. This table shows that the large majority of documents by all systems were indexed in the range of 7 - 12 minutes, and also emphasises again the higher efficiency of Uniterm at the shortest indexing time.

Time	UDC		ALPHA.		FACET		UNITERM	
Minutes	a	b	a	b	a	b	a	b
1 - 3	34	56%	92	67%	22	68%	36	87%
4 - 6	13	69%	91	89%	35	80%	40	85%
7 - 12	356	80%	237	74%	115	80%	239	91%
13 - 18	80	73%	90	77%	92	83%	47	79%
19 - 24	20	75%	-	-	42	74%	11	64%
25 +	62	79%	-	-	7	86%	11	100%

TABLE 8.7

EFFICIENCY AT VARYING INDEXING TIMES FOR SUPPLEMENTARY INDEXING

(a) Number of documents indexed in group

(b) % Successful retrievals