As a postscript, it might be stated that the machine times given are representative of what we were able to obtain when the various pieces of equipment were working satisfactorily. Unfortunately for us the equipment which we used was old and had only been used intermittently for demonstration purposes during the last few years, and we had numerous breakdowns, with the result that the time spent on the operation was probably well over 500 hours. However, we have no reason to think other than that with properly maintained equipment, the figures which are given would be quite practical.

CHAPTER 6

STATISTICAL DETAILS

In Tables 4 - 9 are given various statistical details of the indexing. Some of these may have little value at the present stage, but will become significant in relation to the test results.

The first set of tables gives detailed figures of the postings for each group of 100 documents during the indexing of the final 6,000 documents. As was to be expected, there was a regular falling off in the number of postings with indexing times; in addition we have, as was hoped, a variation between the indexers. While it would obviously be incorrect to suggest that there is a correlation between the standard of indexing and the number of entries required for each document, yet we hope that in the testing we shall be able to ascertain whether over-indexing has an equally bad effect as under-indexing.

Table 5 compares the indexing done during the first two sub-

programmes with the figures from Table 4; the general increase during the whole programme is presumably an indication of the indexers growing familiarity with the systems.

Table 6 shows the variation in postings between research reports and articles in journals, while Table 7 gives data on the postings for individual series and journals. Table 8 gives figures for the new headings originated during the final batch of 6,000 documents. Throughout the whole programme, the number of headings or notations used was as follows :-

U.D.C.	4052	headings in alphabetical index
	2350	different notational elements
Alphabetical	2864	main headings
	1560	"see" references
	592	subheadings
Facet	1686	notational elements
Uniterm	3174	terms

(including 607 Proper names and 267 numbers)

An interesting table (No. 9) concerns the compilation of the chain index to the facet classification. As has already been stated, in theory a card is required for every notational element used in indexing a document. In practice certain elements and combination of elements will be repeated, so that on the second and subsequent occasions when it is used, no card is required for insertion in the chain index. The result of this is that in theory there should come a time when all possible permutations will have been used, and therefore no further cards will require to be entered in the chain index.

Document Numbers	U.D.C.		ALPHABETICAL			FACET		UNITERM
	Total	Elements	Total	Main Headings	Sub Headings	Total	Elements	
12001 - 100	635	339	385	236	210	228	918	1181
12101 - 200	584	340	324	210	194	232	877	1192
12201 - 300	565	325	352	203	194	232	905	1156
12301 - 400	512	303	310	221	233	184	713	1294
12401 - 500	507	311	339	227	252	210	818	868
12501 - 600	420	270	300	227	208	192	654	876
12601 - 700	298	251	271	209	202	173	559	899
12701 - 800	319	243	283	206	217	150	503	1042
12801 - 900	382	266	261	183	210	150	525	961
12901 -13000	438	297	281	209	225	159	602	922
13001 -13100	388	274	274	205	219	179	679	917
13101 -13200	363	272	278	210	228	156	577	1044
13201 -13300	376	250	264	215	172	156	496	857
13301 -13400	336	257	265	209	184	148	505	782
13401 -13500	284	220	219	184	124	149	412	746
13501 -13600	339	255	242	193	154	146	483	994
13601 -13700	283	219	220	165	180	141	396	720
13701 -13800	304	258	245	187	213	143	475	745
13801 -13900	315	237	240	174	195	151	479	749
13901 -14000	240	212	199	169	160	120	369	842

TABLE 4A DATA ON POSTINGS FOR DOCUMENT GROUPS DURING FINAL SUB-PROGRAMME (DOCUMENTS 12,001 - 18,000) INDEXER A

Note: The columns "Total" represent the number of cards required by the indexer to be put in the catalogues. For U.D.C. and Facet the column "Elements" represents the number of different notational elements which were actually used. In Alphabetical, "Main Headings" and "Sub-Headings" represent the different headings which were actually used.

Document Numbers	U.D.C.		ALPHABETICAL			FACET		UNITERM
	Total	Elements	Total	Main Headings	Sub Headings	Total	Elements	
14001 - 100	1118	646	512	362	219	149	645	1141
14101 - 200	468	351	1030	601	422	117	509	988
14201 - 300	425	345	314	253	207	218	1294	1001
14301 - 400	338	303	264	201	199	113	460	2451
14401 - 500	661	437	313	231	224	143	489	882
14501 - 600	412	328	415	309	285	131	533	845
14601 - 700	417	336	274	207	209	241	1091	952
14701 - 800	376	329	242	200	200	147	518	2012
14801 - 900	543	401	314	235	180	117	482	815
14901 -15000	308	294	502	323	203	117	434	843
15001 - 100	302	282	251	185	170	230	925	781
15101 - 200	281	277	179	159	127	105	400	1654
15201 - 300	339	313	191	161	123	111	342	634
15301 - 400	229	241	289	248	162	118	353	664
15401 - 500	284	293	201	162	145	130	648	744
15501 - 600	257	260	190	172	124	117	278	1108
15601 - 700	284	279	168	159	140	107	337	740
15701 - 800	220	222	230	190	158	114	220	681
15801 - 900	241	245	162	159	131	128	415	629
15901 -16000	202	213	164	141	108	109	302	900

TABLE 4B DATA ON POSTINGS FOR DOCUMENT GROUPS DURING FINAL SUB - PROGRAMME (DOCUMENTS 12,001-18,000) INDEXER B

Note: The columns "Total" represent the number of cards required by the indexer to be put in the catalogues. For U.D.C. and Facet the column "Elements" represents the number of different notational elements which were actually used. In Alphabetical, "Main Headings" and "Sub-Headings" represent the different headings which were actually used.

Document Numbers	U.D.C.		ALPHABETICAL			FACET		UNITERM
	Total	Elements	Total	Main Headings	Sub Headings	Total	Elements	
16001 - 100	765	541	565	399	295	173	725	1113
16101 - 200	668	509	607	401	321	203	770	1045
16201 - 300	614	448	487	350	267	242	935	1052
16301 - 400	712	499	610	373	307	223	842	1264
16401 - 500	760	483	548	373	275	198	795	991
16501 - 600	645	444	528	346	304	215	832	972
16601 - 700	556	432	428	319	270	204	842	927
16701 - 800	518	409	436	306	241	172	685	955
16801 - 900	482	396	389	304	215	158	538	761
16901 -17000	403	345	359	265	233	152	455	695
17001 - 100	398	344	327	231	183	165	570	698
17101 - 200	481	379	419	296	220	141	515	930
17201 - 300	370	311	275	211	185	113	387	638
17301 - 400	376	328	310	235	195	119	421	683
17401 - 500	390	329	319	235	187	118	448	696
17501 - 600	393	319	291	234	161	127	426	734
17601 - 700	235	242	216	173	128	110	314	533
17701 - 800	280	259	228	192	162	110	335	573
17801 - 900	282	260	222	188	136	104	367	589
17901 -18000	277	252	217	187	140	102	335	616

TABLE 4C DATA ON POSTINGS FOR DOCUMENT GROUPS DURING FINAL SUB-PROGRAMME (DOCUMENTS 12,001 - 18,000) INDEXER C

Note: The columns "Total" represent the number of cards required by the indexer to be put in the catalogues. For U.D.C. and Facet the column "Elements" represents the number of different notational elements which were actually used. In Alphabetical, "Main Headings" and "Sub-Headings" represent the different headings which were actually used.

TOTALS FOR DIFFERENT TIME ALLOWANCES

Indexer A	U.D.C.		ALPHABETICAL			FACET		UNITERM
	Total	Elements	Total	Main Headings	Sub Headings	Total	Elements	
16 minutes	2317	1307	1371	870	831	876	3413	4813
12 minutes	1544	1075	1193	869	879	725	2534	3685
8 minutes	1571	1109	1094	807	882	644	2383	3844
4 minutes	1335	982	990	801	634	599	1896	3379
2 minutes	1142	926	904	695	748	555	1719	3056
Indexer B								
16 minutes	2349	1645	2120	1417	1047	597	2908	5581
12 minutes	1866	1430	1244	947	918	662	2631	4691
8 minutes	1434	1254	1246	902	680	569	2241	4093
4 minutes	1109	1107	871	783	554	576	1621	3150
2 minutes	947	959	724	649	537	451	1274	2950
Indexer C								
16 minutes	2763	1997	2069	1523	1190	841	3278	4464
12 minutes	2479	1768	1940	1344	1090	789	3154	3845
8 minutes	1764	1464	1494	1096	851	616	2078	3084
4 minutes	1529	1287	1195	915	728	477	1682	2751
2 minutes	1074	1013	883	740	566	426	1351	2311

TABLE 4D TOTAL POSTINGS FOR DIFFERENT TIME ALLOWANCES DURING FINAL BATCH

Note: The columns "Total" represent the number of cards required by the indexer to be put in the catalogues. For U.D.C. and Facet the column "Elements" represents the number of different notational elements which were actually used. In Alphabetical, "Main Headings" and "Sub-Headings" represent the different headings which were actually used.

		U.D.C.		ALPHABETICAL			FACET		UNITERM
		Total	Elements	Total	Main Headings	Sub-Headings	Total	Elements	
Indexer A									
16	a	427	352	301	248	153	101	554	915
	b	435	255	249	194	162	132	434	919
	c	579	327	343	217	208	219	853	1203
8	a	310	252	226	178	138	100	380	858
	b	405	255	268	203	175	157	496	939
	c	393	277	273	202	220	161	596	961
2	a	224	189	168	141	98	105	276	651
	b	249	205	232	180	158	132	391	831
	c	285	231	226	174	187	139	430	764
Indexer B									
16	a								
	b	293	296	365	322	190	105	606	1106
	c	587	411	530	354	262	149	727	1395
8	a								
	b	284	286	296	261	140	116	425	954
	c	358	313	314	225	170	142	560	1023
2	a								
	b	177	184	179	221	79	102	284	545
	c	237	240	181	162	134	114	318	737
Indexer C									
16	a	306	269	283	270	98	102	519	608
	b	488	412	436	293	214	130	567	834
	c	691	499	517	381	297	210	809	1116
8	a	373	294	288	247	130	101	384	595
	b	376	339	350	260	188	112	472	727
	c	441	366	373	274	213	154	519	771
2	a	215	130	203	159	115	101	278	454
	b	196	208	193	159	137	102	298	542
	c	268	253	221	185	141	106	338	573

TABLE 5A AVERAGE INDIVIDUAL POSTINGS PER HUNDRED DOCUMENTS FOR 16 MINUTES, 8 MINUTES AND 2 MINUTES FOR THREE SUB-PROGRAMMES

a = Documents Nos. 1 - 6000, b = Nos. 6001 - 12000, c = Nos. 12001 - 18000

N.B. Indexer B did not index any documents in sub-programme 1 - 6000

		U.D.C.		ALPHABETICAL			FACET		UNITERM
		Total	Elements	Total	Main Headings	Sub-Headings	Total	Elements	
16	a	366	310	292	259	125	101	536	761
	b	405	321	350	270	189	122	536	953
	c	619	412	463	317	256	193	796	1238
8	a	341	273	257	212	134	100	382	726
	b	355	293	305	241	168	128	464	873
	c	397	318	320	227	201	152	592	918
2	a	219	159	185	150	106	103	277	552
	b	243	199	201	187	135	112	324	637
	c	263	241	209	174	154	120	362	691

TABLE 5B AVERAGE COMBINED POSTINGS PER HUNDRED DOCUMENTS FOR 16 MINUTES, 8 MINUTES AND 2 MINUTES FOR THREE SUB-PROGRAMMES

a = documents 1 - 6000

b = documents 6001 - 12000

c = documents 12001 - 18000

Time in minutes	U.D.C.		ALPHABETICAL		FACET		UNITERM	
	Report	Jnl.	Report	Jnl.	Report	Jnl.	Report	Jnl.
16	9.1	7.6	5.0	4.7	1.8	1.7	11.8	11.0
12	5.5	4.4	4.5	3.8	1.9	1.7	9.5	8.4
8	4.7	3.2	3.0	2.7	1.9	1.8	8.6	7.3
4	3.7	2.8	2.6	2.2	1.3	1.3	10.4	8.5
2	2.8	2.3	2.2	2.2	1.2	1.2	7.2	6.0
Average	5.4	4.1	3.5	3.2	1.6	1.5	9.5	8.2

TABLE 6 AVERAGE POSTINGS PER DOCUMENT FOR REPORTS AND JOURNAL ARTICLES

	U.D.C.	ALPHABETICAL	FACET	UNITERM
Royal Aircraft Est. Reports and Notes	5.5	3.5	1.7	9.0
National Advisory Committee for Aeronautics. Notes and Memoranda	6.1	3.8	1.6	10.3
Flight & Aeroplane	1.5	1.6	1.4	6.3
Aircraft Engineering	5.5	4.5	2.0	9.3
Metallurgica & Metal Progress	7.3	5.3	2.3	12.0
Royal Aeronautical Society Journal	3.6	3.1	1.8	6.8
Journal Aeronautical Sciences	4.1	3.2	1.5	8.8

TABLE 7 AVERAGE POSTINGS BY SERIES AND JOURNALS

	U.D.C.			ALPHA-BETICAL			UNITERM		
Indexer:-	A	B	C	A	B	C	A	B	C
Indexing Time									
16	31	22	47	30	19	29	16	34	6
8	16	2	24	25	14	14	12	10	4
2	4	6	19	5	13	9	10	10	1
4	9	3	16	14	1	11	11	2	8
12	13	14	33	21	19	6	22	17	5
Individual Totals	76	47	139	95	66	69	71	73	24
Group Totals	265			230			168		

TABLE 8 NEW HEADINGS ORIGINATED FOR GROUPS OF 100 DOCUMENTS IN FINAL BATCH AT VARIOUS INDEXING TIMES

No valid information was available as to how long it would take to reach that position or alternatively whether there was any likelihood of the stage being approached.

Clearly the more compact the subject field, the greater the chance that the possible permutations will be repeated. Further, where the notation is made up of few separate elements, the notation is more likely to be repeated than in those cases where the notation is long and complex. As will be seen from Table 9, the figure for cards required dropped fairly quickly, but appeared to level out at the figure of approximately 50% and it seems that continued indexing would not bring it very far below the figure.

A claim made for the facet classification is the economy in total entries which results from a fixed order with the chain index. For the three card catalogues during the final six thousand documents, the following figures show the number of cards actually inserted.

U.D.C.		Alphabetical	Facet	
Classified catalogue	Alphabetical Index		Classified Catalogue	Chain Index
25223	1736	19338	9511	18083
Total: 26,959		19,338	27,594	

It would appear that in our subject field it is doubtful if the position would ever be reached where the total number of cards required for the facet catalogue was appreciably less than those considered necessary for the other systems.

We did not specify any term in facet as an "unsought" term, although we were aware that certain terms, such as Tests, Wind Tunnel Tests, Flight Tests, Design, Analysis, Calculation and Measurement, would in

No. of documents	Possible No. of chain index cards	Cards Inserted	Cards Duplicated	% of possible cards inserted	Duplicate cards with three or more elements			
					3	4	5	6
1200	4111	3094	1017	75.2	21	2	-	-
1200	3711	2228	1483	60.0	51	7	1	1
1200	5101	2990	2111	58.8	84	17	4	1
1200	5271	2937	2334	55.8	135	46	10	3
1200	4456	2385	2071	53.0	133	34	7	1
1200	4855	2379	2476	49.0	154	33	8	1
1200	6943	3271	3672	47.3	263	72	18	3
1200	7709	4277	3432	55.5	292	78	20	4
1200	7309	3865	3444	52.8	302	81	22	3
1200	5268	2719	2549	51.6	253	71	20	1
1200	6946	3211	3735	46.2	382	113	23	2
1200	8121	4011	4110	49.5	425	118	26	4

TABLE 9 CARDS INSERTED IN CHAIN INDEX OF FACET CATALOGUE DURING WHOLE PROGRAMME

most cases be unsought terms as far as the chain index was concerned. This matter will be investigated in the test programme, but if it had been decided not to prepare chain index entry for any of these terms the number of cards inserted would have been reduced from 18,083 (as shown above) to 16,475. On the other hand, as discussed in Chapter 4, we did not do full chain indexing up to the containing head. If this had been done, it is estimated that the number of inserted cards would have risen to over 30,000.

CHAPTER 7

SUPPLEMENTARY INDEXING

For the testing programme we estimated that we should require 1600 questions which were to be based on documents in the collection, and we sought the co-operation of a large number of organisations and individuals. Discussion of the questions will be given in a later report dealing with the test programme, and it is sufficient to say here that we obtained the required number of questions.

We decided that at the same time we would invite people to index documents for us. The original purpose for doing this was so that we might have a comparison with which the standard of indexing of the project staff could be checked. The following is an extract from the notice which we sent to those whom we thought might be interested enough to help.

> "Assistance of two kinds is requested from individuals or organisations who are interested in this project. Firstly, we wish to have a number of the project documents indexed by persons or groups of persons who are familiar with the subject content,