Chapter 4.

RESULTS

4.1 Problem Statements - general characteristics

We collected problem statements from 27 interviewees, which ranged from 490 to 66 tokens in length. Our initial analysis of these statements was to remove non-significant words from the text of the interviews. As a measure of the 'content density' we computed the ratio of tokens before and after this procedure for each interview, and as a measure of 'redundancy', the type-token ration after removing non-significant words. These data are displayed in Table 1, from which one can note that the problem statements are all quite similar in these characteristics, regardless of length. Using the results of the text analysis program, Table 2 indicates the maximum and minimum (of the top 40 associates) association strengths for each oral problem statement, and the number of types (after removing non-significant words). Maximum association strengths ranged from 1481 to 166, minimum from 132 to 25, and number of types from 99 to 21.

We were concerned to see if there were correlations between association strengths and text characteristics, and if there were differences among the oral problem statements, written problem statements and abstracts. For the oral problem statements, the values of r (the product-moment correlation) calculated were:

1. Highest association strength vs.
   number of types
   r = 0.5676
2. Range of association strengths vs.
   number of types
   r = 0.5547

Both these values are significant at the 1% level (two-tailed t test).

In order to see if there were consistent differences between general text characteristics of oral and written problem statements, we performed the same analysis on eight written problem statements. These results are displayed in Tables 3 and 4, which show a mean 'content density' of 2.05, as compared to 2.7 for oral problem statements.

The values for r, the product-moment correlation, for the written problem statements were:

1. Highest association strength vs.
   number of types
   r = 0.377

Interview No.	Pre / Post Tokens	Post Types / Tokens
1	2.49	0.48
2	2.78	0.67
3	3.09	0.78
4	3.06	0.66
5	2.40	0.78
6	2.97	0.74
7	2.00	0.69
8	2.27	0.89
9	3.76	0.71
10	2.89	0.67
11	2.13	0.45
12	3.06	0.62
13	2.68	0.81
14	2.69	0.53
15	2.30	0.61
16	2.63	0.73
17	2.59	0.37
18	2.57	0.60
19	2.20	0.70
20	2.47	0.64
21	2.27	0.70
22	2.76	0.76
23	3.78	0.72
24	2.78	0.66
25	2.87	0.77
26	2.39	0.69
27	2.98	0.82
Mean	2.7	0.69
Variance (S^2)	0.3	0.01

Pre - Before Identification of Significant Words

Post- After Identification of Significant Words

Table 1. Token-token and type-token ratios for oral problem statements (from Brooks, 1978).

Interview No.	Highest Association Strength	Lowest Association Strength	Range of Strengths	No. of types
1	879	132	47	95
2	316	50	266	68
3	198	33	165	61
4	257	50	207	47
5	258	33	225	55
6	430	42	372	56
7	631	33	598	31
8	273	33	240	33
9	497	58	439	60
10	297	50	247	37
11	464	26	438	64
12	924	75	849	49
13	273	58	215	43
14	1481	52	1349	87
15	1114	83	1031	84
16	183	25	158	37
17	166	25	141	30
18	489	48	431	80
19	249	25	224	21
20	390	100	290	99
21	297	56	231	32
22	265	33	232	34
23	264	33	231	26
24	447	50	397	53
25	297	58	239	46
26	290	50	240	42
27	264	66	198	46

Table 2. Association strengths and number of types for oral problem statements (from Brooks, 1978).

Script No.	$\frac{\text{Pre}}{\text{Post}}$ Tokens	Post $\frac{\text{Types}}{\text{Tokens}}$
30	2.08	0.7
31	2.29	0.85
32	1.84	0.68
33	2.00	0.75
34	1.86	0.71
35	1.55	0.71
36	2.09	0.76
37	2.68	0.68
Mean	2.05	0.73
Variance (S^2)	0.113	0.01

Table 3. Token-token and type-token ratios for written problem statements (from Brooks, 1978)

Script No.	Highest Association Strength	Lowest Association Strength	Range of Association Strengths	No. of Types
30	198	33	165	42
31	141	33	108	35
32	183	33	150	35
33	100	25	75	24
34	657	33	624	42
35	1242	33	1209	39
36	274	33	241	34
37	258	25	233	15

Table 4. Association strengths and number of types for written problem statements, (from Brooks, 1978).

2. Range of association strengths vs. number of types
   r = 0.384

Neither of these values is significant, even when recomputed after having removed number 35 from the data because its maximum association strength is so much higher. Thus, there appears to be no relation between this text parameter and association strength values.

Finally, in order to indicate the subject spread of the problem statements, we classed them into the five broad categories indicated in Table 5. The social sciences are well represented in our sample, as is medicine, with perhaps some under representation of technology and the natural sciences. Nevertheless, the spread, given sample size, is reasonably broad.

4.2 Problem Statements - evaluation *

The point of the surveys of users and authors was to see whether the analyses of the problem statements and abstracts were in general accord with the originators' own perceptions of their information needs or of the ideas they were attempting to communicate; and, if there were disparities, then to see if there might be suggestions for improvement. Of course, for retrieval purposes it may not be necessary for the representations to be congruent with the originators' ideas about them, but as a first method of evaluation the technique seemed reasonable. If the subjects were unanimous in their disapproval of the representations, then we could be fairly sure that we should probably try something else.

We wished to determine in evaluating the problem statement representation:

1. how accurately, in the interviewee's opinion, the two formats described her/his ASK at the time of the interview; and
2. how the two formats compared with one another.

Response to the survey was good, 63% of the group (N = 27) returning completed questionnaires. Table 6 is a summary of replies to the Association Map questionnaire, Table 7 to the Association Clusters questionnaire, and Table 8 to the comparative questionnaire.

From these tables it is evident that the analysis, presented in the Association Map format, provided a generally adequate representation of the information needs of the interviewees. The major criticism of the analysis is that some concepts were too weakly associated, and this seems

* This section, and section 4.4, are based on Brooks, Oddy and Belkin (1979)

	Psychology /Education /Sociology /Linguistics	Medicine	Agriculture	Information Science	Biology /Chemistry /Bio-Chemistry
	1	3	9	14	7
	2	4	10	15	17
	6	5	11	16	26
	8	19		18	27
	12	20			
	13	21			
	22	23			
	24	25			
Totals	8	8	3	4	4

Table 5. Subject areas of interviewees by interview number (from Brooks, 1978).

QUESTION	% YES	% NO	NO RESP. (N=15)
1. ACCURATE REFLECTION?	73.3	20.0	6.6%
2.(A) TOO STRONG?	46.6	53.4	-
(B) TOO WEAK?	86.6	13.4	
3. CONCEPTS MISSING?	46.6	40.0	13.3%

TABLE 6 ASSOCIATION MAP EVALUATION (PROBLEM STATEMENT)

QUESTION	% YES	% NO	NO RESP. (N=15)
1.(A) CONCEPTS WHICH SHOULD NOT BE TOGETHER?	46.6	46.6	6.6
(B) CONCEPTS WHICH SHOULD BE TOGETHER?	73.3	20.0	6.6
2.(A) GROUPS AT TOO HIGH AN ASSOCIATION LEVEL?	60.0	40.0	
(B) GROUPS AT TOO LOW AN ASSOCIATION LEVEL?	60.0	40.0	
3.(A) GROUPS NOT CLOSELY ENOUGH LINKED?	80.0	20.0	
(B) GROUPS TOO CLOSELY LINKED?	46.6	53.4	

TABLE 7 ASSOCIATION CLUSTER EVALUATION (PROBLEM STATEMENT)

QUESTION	% YES	% NO	NO RESP.	
1. ANY PREFERENCE?	73.0	26.6		N=15
2. PREFER ASSOCIATION MAP	81.8			N=11
3. PREFER ASSOCIATION CLUSTER	18.2			N=11
4. IF NO PREFERENCE, WERE BOTH UNSUCCESSFUL?	75.0		25.0%	N=4

TABLE 8 FORMAT COMPARISON

to be its single identifiable consistent problem, in the judgement of the interviewees. The Association Cluster format was judged inferior to the Association Map.

4.3 Abstracts - general characteristics

We collected 31 abstracts from the library and information science literature, two of which were of the same article, one a very long synopsis, the other a short abstract.

In order to compare abstracts with problem statements, we performed the same analyses for 'content density' and 'redundancy', and also determined maximum and minimum association strengths and number of significant types for each abstract. These data are displayed in Tables 9 and 10. The values for r for the abstracts were:

1. Number of types vs. highest association strength
   r = 0.54315
2. Number of types vs. range of association strengths
   r = 0.5191

Both of these values are significant at the 1% level (two-tailed t test), so we conclude that association strengths are related to parameters of the abstracts.

4.4 Abstracts - evaluation

The goals of the abstract evaluation were to see:

1. how accurately the analysis, presented in the Association Map format, represented the interrelation of concepts in the mind of the author at the time of writing; and
2. if there were any regularities in deficiencies of the representation.

We chose to use only the Association Map format because of the general dissatisfaction with the Cluster format among the interviewees who responded to their questionnaire.

The response rate to this survey was gratifyingly high: 90% (N = 30). The results are summarized in Table 11. There are, unfortunately, some difficulties in interpreting these results (see Section 6.1), but it appears that although the general representation method is judged reasonable, there are some severe problems in its specific implementation. Thus, although only about 30% of the respondents thought that the representation was actually bad, 63% thought that some concepts were actually omitted, and most striking, 96% (all but one) thought that at least some concepts were too weakly connected.

Abstract No.	Pre Tokens / Post	Post Types / Tokens
1	1.91	0.536
2	1.895	0.789
3	1.828	0.737
4	2.32	0.681
5	2.16	0.5816
6	1.95	0.66
7	1.99	0.656
8	2.026	0.608
9	1.86	0.765
10	2.058	0.662
11	2.06	0.62
12	1.987	0.73
13	2.03	0.609
14	2.15	0.814
15	1.94	0.597
16	1.876	0.609
17	2.218	0.705
19	1.74	0.717
20	1.66	0.696
21	1.977	0.699
22	2.12	0.405
23	1.776	0.606
24	1.84	0.737
25	2.4	0.745
26	1.646	0.747
27	2.02	0.637
28	2.078	0.429
29	1.78	0.651
30	1.948	0.724
31	2.03	0.644
32	1.87	0.654
Mean	1.97	0.658
Variance	0.03	0.008

Table 9. Token-token and type-token ratios for abstracts.

Abstract No.	Highest Association Strength	Lowest Association Strength	Range of Strengths	No. of Types
1	266	99	167	30
2	291	58	233	60
3	373	66	307	73
4	257	83	174	32
5	1071	91	980	57
6	481	133	348	70
7	332	83	249	59
8	447	66	381	93
9	290	58	232	62
10	341	83	258	45
11	447	100	347	80
12	472	83	389	57
13	415	108	307	78
14	116	58	58	48
15	539	83	456	40
16	132	50	82	78
17	300	58	292	55
19	150	66	84	33
20	283	58	225	39
21	315	75	140	93
22	1053	108	945	242
23	1011	190	821	57
24	166	58	108	28
25	399	83	316	79
26	274	66	208	74
27	422	75	347	79
28	1058	141	917	110
29	472	83	389	84
30	158	66	92	42
31	704	100	604	56
32	365	91	274	61

Table 10. Association strengths and number of types for abstracts.

QUESTION	% YES	% NO	% INTERM.	% NO RESP.	
1. ACCURATE REFLECTION?	48.0	29.6	22.0		N=30
2.(A) CONCEPTS TOO STRONGLY CONNECTED?	63.0	37.0			N=30
(B) CONCEPTS TOO WEAKLY CONNECTED?	96.3	3.7			N=30
3. CONCEPTS OMITTED?	88.9	11.1			N=30
4. IF NO OR 'INTERM' TO No.1, WAS ABSTRACT ACCURATE?	64.3	7.1	21.4	7.1	N=14

TABLE 11 ABSTRACT REPRESENTATION EVALUATION