CHAPTER 6

DISCUSSION

6.1 Design Problems

There are several problems in interpreting our data. The major difficulty
lies in the questionnaire design, and has affected the results concerning
evaluation of the text analysis. It is especially evident in the results
for the abstract evaluation. From the responses to that questionnaire
(See Appendix B) it is clear that we should have used a rating scale
rather than 'yes/no' binary choice for the general evaluative question.
A number of respondents expressed difficulty in answering that question,
and answers to that question were not always consistent with answers
to the other questions. Although this was not made evident in the problem
statement evaluation, that may be only because the respondents to the
abstract evaluation were professionally familiar with the issues, and
therefore more disposed to comment and criticize. So we might assume
that for both groups, the judgement of general adequacy of the analysis
has been constrained by the response choice. This also seems to have
been a difficulty with question 4 of the abstract evaluation, but in
that case is perhaps not quite so critical.

Another problem is that it is difficult to correlate responses to the
questions about strength of association with those asking if any concepts
were omitted. It is possible that some omitted concepts might have been
those which were also too weakly related. Although we asked for examples
of each case, there is no certain way to determine whether such a
relationship holds. And finally, we did not stress, in our explanation,
that the associations in the representation formats were only a small
number of the total associations derived by the analysis, representing
only the most important concepts and relationships. This may have
coloured responses to questions concerning weak associations and omitted
concepts, and also seems to have influenced response to question 1,
negatively.

Another general difficulty ties in the representations of problem statements,
where we arbitrarily chose the top 40 associates for inclusion in the
association map representation. This number was chosen for ease of construc-
tion and interpretation of the representation, but unfortunately resulted
in arbitrarily splitting groups of associates all with the same association
strength. We realized this difficulty by the time the abstracts were
evaluated, and so used a flexible cut-off level with them in order to
include all associates at the lowest association strength, aiming for

about 40 associates. This problem may have affected response to the evaluation of problem statement representations, and cut-off level certainly needs further investigation (but see Section 6.4).

6.2 Text characteristics

The general text characteristics with which we were concerned, i.e. the ratio of all tokens to significant tokens, and the type-token ratio, followed expected patterns. Oral problem statements had the highest token-token ratio, abstracts the lowest and written problem statements were intermediate, supporting the general finding (e.g. Mann 1949; Moses 1959; Portnoy, 1973) that oral narrative has more non-content words than written narrative. From our results, we might conclude that the written problem statements we examined were not so fully thought out as the abstracts. There is, however, one striking difference between the three categories of text; that is, the correlation between number of types and association strength. In this case both oral problem statements and abstracts showed significant correlations, while the written problem statements showed none. This may perhaps be due to the style of the written statements, which appeared to be strongly affected by previous experience with information systems. The users tended to put their queries together in a rather boolean form, which led to high association strengths irrespective of number of types. Whether this difference would be significant in system design is not, however, clear, since the text analyses of written and oral problem statements were quite similar.

6.3 Evaluation of representations

Before discussing the evaluation results themselves, we would like to comment briefly upon the response rates to our surveys, which seem exceptionally high. We decided from the onset of the project to pay our respondents for taking part in the evaluation and we feel that this has had a great influence upon the response rate. For the interviewees, we included a cheque for only £2.00, having underestimated the amount of work the evaluation would require. We received back eight of these cheques; from this we conclude that our having made even only a gesture toward repaying the respondents for their effort was sufficient to affect response rate. We increased the gesture to £5.00 for the abstract evaluation. Although there are probably other reasons for our very high response rate (90%), such as professional involvement in the subject of the survey, we cannot help but think that recognition of effort in the form of payment was a substantial factor in the response. There were several comments to this effect from the respondents, only one of

whom returned a cheque. We sent no reminders in either survey, so this
sort of payment seems to us a relatively inexpensive way to achieve good
response rates (at least in small-scale surveys).

The results of the evaluations, despite the difficulties mentioned in
Section 6.1, seem to us encouraging. In both groups of respondents,
30% or less thought that the representation was actually bad, and in
some of those cases further comments contradicted that opinion. From
these data, we conclude that the basic analytic technique, word co-occur-
rence analysis, is in principle reasonable for our purposes. The results
of the survey of authors also seem to indicate that abstracts are
reasonable document surrogates in our context, and that the length of the
abstract is not a significant parameter, at least within the limits in
our sample. And the results of the survey of interviewees indicates that
our interview technique, combined with the text analysis, is a reasonable
means for collecting problem statements that will be useful in an ASK
based IR system. Despite these positive results of the survey, the
data concerning strength of association and omitted concepts indicate
that our text analysis techniques need modification.

6.4 Text analysis

The results of the evaluations of both problem statement and abstract
representations indicate that over-strong associations are a problem, but
that the major difficulty in text analysis lies in associations at too low
a level. These may reflect a misunderstanding of the representation, as
mentioned in Section 6.1, but they seem too consistent and too positive to
be attributed only to this cause. Most of the examples of too weakly
associated concepts involve words that occurred only once or twice in
the text, sometimes in conjunction with more frequent words, but often not.
At times these words appear (in the abstracts) only in the document title,
which we treated as a separate paragraph. Possible means of adjusting our
analysis are to include the title as part of every paragraph, and to take
account of collection frequency (when we have a large enough collection)
to weight some of the content words. We also need to adjust cut-off
techniques in some way, to include more than just the top 40 associates,
but fewer than all of the associates in the text representation. We have
done some preliminary investigations on this problem, and it appears that
using all associates until each type is represented is a reasonable
and useful cut-off rule.

The data from the text analysis as it now stands are rather complex,
and probably need to be reduced before they can be useful for retrieval

purposes (as, for instance, suggested in Section 5.2). We have done this to some extent in the association map representations by choosing three levels of association strength, but the techniques for choosing levels need to be refined.

The classification of problem statements based on the text analysis appears to be reasonably successful, given its aims. It is interesting to note that there are few queries in our sample which can be described as well-defined. The one in class A0 stands out in this respect, from both the subjective and computational points of view (it has a connectivity score of zero). We believe that conventional best-match retrieval systems operate under the assumptions that queries are well-defined. Our results appear to belie that assumption, and it is encouraging that we can come to this conclusion by computational and algorithmic means. Nevertheless the classification should be modified in order to take more account of levels of association.