

City Research Online

City St George's, University of London

Citation: Yates-Mercer, P. A. (1973). Investigation of relational indexing for selective dissemination of information. (Unpublished Doctoral thesis, The City University)

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/38440/

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

INVESTIGATION OF

RELATIONAL INDEXING

FOR SELECTIVE DISSEMINATION OF INFORMATION

Penelope Ann Yates - Mercer

Thesis submitted for the degree of Doctor of Philosophy

THE CITY UNIVERSITY LIBRARY,
ST. JOHN STREET, LONDON, E.C.1.

The City University

Centre for Information Science

July 1973

CONTENTS

		Page
	Table of Figures	iv
	Tables	vi
	ABSTRACT	viii
1.	INTRODUCTION	1
2.	METHODOLOGY	11
2.1	Testing methodology	12
2.2	Data base	14
2.3	"Users"	17
2.4	Profile collection	20
2.5	Profile construction	22
2.6	Matching	25
2.7	User assessment	28
2.8	Evaluation measures	30
3.	RESULTS AND DISCUSSION	37
3.1	Relational indexing	37
3.1.1	The relations	37
3.1.2	Number of relations associated with one concept	39
3.1.3	Three dimensional representation	42
3.1.4	Indexing and preparation times	45
3.1.5	Exhaustivity of indexing	46
3.1.6	Breadth of indexing	48
3.1.7	Generic browsing	49
3.1.8	Conceptual browsing	51
3.1.9	Condensations	52
3.1.10	Related browsing	58
3.1.11	Analysis of selection strategies	64

		Page
3.2	Evaluation	68
3.2.1	Primary parameters	68
3.2.2	Calculated parameters	69
3.2.3	Effect of parameter values	70
3.2.4	Variation over the six months	72
3.2.5	Cumulation of results	73
3.2.6	Overall performance over six months	74
3.2.7	The recall - fallout graph	80
3.2.8	The "generality" effect	85
3.2.9	Distribution of results	89
3.2.10	The value of N	94
3.2.11	An "overall percentage"	97
3.2.12	The 0/0 situation	100
3.2.13	Categories of relevance	101
3.2.14	Size of data base for testing	105
3.2.15	Number of questions required	106
3.2.16	Error analysis	109
3.2.17	Effects of searching strategies	116
3.2.18	Consistency of indexing	121
	(Graphs relating to Section 3.2)	126
3.3	SDI Systems	167
3.3.1	Profile collection and construction	167
3.3.2	Analysis of profiles	171
3.3.3	Mode of operation	178
3.3.4	Organizational groups	185
3.3.5	Profile modification	187
3.3.6	Comparison of two similar profiles	190
3.3.7	Size of user selections	192
3.3.8	Relevance	195
4.	GENERAL DISCUSSION	204
4.1	Methodology	204
4.2	User needs	211
4.3	Users' comments	213
5.	CONCLUSIONS	218

		Page
	ACKNOWLEDGEMENTS	224
	REFERENCES	226
	APPENDICES	
I	Form on which users noted numbers of abstracts of interest	237
II	Tables of results (arranged by set)	238
III	Examples of abstracts with their relationally indexed displays	258
IV	Examples of the browsing techniques	267

Table of Figures

		Page
1.	The SDI communication chain	2
2.	Guilford's cubic "structure of intellect" model	4
3.	The relational operators	5
4.	Typical card index entries for one abstract	18
5.	2 x 2 contingency table	31
6.	The difference between generic and related browsing	60
7.	Recall-precision and recall-fallout diagrams (set 071)	126
8.	" " " " " " (set 072)	127
9.	" " " " " " (set 073)	128
10.	" " " " " " (set 083)	129
11.	" " " " " " (set 093)	130
12.	" " " " " " (set 103)	131
13.	" " " " " " (set 113)	132
14.	" " " " " " (set 123)	133
15.	Monthly average and cumulative mean of recall (set XY3)	134
16.	" " " " " of precision (set XY3)	134
17.	" " " " " of fallout (set XY3)	135
18.	Recall-precision and recall-fallout diagrams (monthly average and cumulative mean)	136
19.	Recall-precision and recall-fallout diagrams (set 712)	137
20.	General recall-fallout diagram, showing Q values	76
21.	Recall-precision diagram showing lines of constant (M x P)	78
22.	Recall-fallout diagram showing lines of constant Q	82
23.	1/M - 1/F graph (set 123)	138
24.	Recall - log fallout graph (set 123)	139
25.	Log recall - log fallout graph (set 123)	140
26.	General recall-fallout diagram	84
27.	Recall-generality graph (set 123)	141
28.	Log recall - log generality graph (set 123)	142
29.	Recall - log generality graph (set 123)	143
30.	1/M - 1/G graph (set 123)	144
31.	Recall - 1/generality graph (set 123)	144
32.	Precision - generality graph (set 123)	145

Page

33. Log precision - log generality graph (set 123) 146
34. Precision - log generality graph (set 123) 147
35. 1/precision - 1/generality graph (set 123) 148
36. Precision - 1/generality graph (set 123) 148
37. Fallout - generality graph (set 123) 149
38. Recall and precision frequency distributions (0.05 intervals) (set 712) 150
39. Recall and precision frequency distributions (0.10 intervals) (set 712) 151
40. Fallout frequency distributions (0.001 and 0.002 intervals) (set 712) 152
41. Recall, precision and fallout distributions (set 121) 153
42. " " " " " (set 122) 154
43. " " " " " (set 123) 155
44. Recall and precision distributions (0.05 intervals) (set 123) 157
45. Q value distribution (set 712) 158
46. (M x P) distribution (set 712) 158
47. Generality distribution (set 712) 159
48. Recall-generality and precision-generality graphs, with second value of N (set 123) 160
49. Fallout-generality graph, with second value of N (set 123) 161
50. Recall-precision diagrams (relevance 1 and relevance 2) (set 123) 162
51. Recall-fallout diagrams (relevance 1 and relevance 2) (set 123) 163
52. Cumulative average of recall (sets 121, 122 and 123) 164
53. Cumulative average of precision (sets 121, 122 and 123) 165
54. Recall-precision diagrams for selection strategies 166
55. Recall-precision and recall-fallout diagrams showing different profile interaction (set 712) 169
56. Model of a general information storage and retrieval system 205

Tables

		Page
1.	Number of abstracts published in five sections of Metals Abstracts in June to December 1970	16
2.	Frequency of occurrence of relations in examples with more than four relations	41
3.	Average times to perform operations in preparing the card index	45
4.	Analysis of the generic browsing strategy	50
5.	Analysis of simple condensations	54
6.	Analysis of double, triple and quadruple condensations	55
7.	Occurrences of single jumpbacks	57
8.	Analysis of related browsing occurrences	63
9.	Monthly analysis of searching strategies	66
10.	Analysis of selection strategy most used by profiles	67
11.	Arithmetic means of results (for relevance (1 + 2))	71
12.	Monthly Q values	79
13.	Correlation of Q and (M x P) (set 712)	81
14.	Comparison of means (set 712)	89
15.	Comparison of means (sets 121, 122 and 123)	91
16.	Arithmetic means for modes of operation (sets 121, 122, 123 and 712)	93
17.	Comparison of averages (set 123)	98
18.	Performance of separate sections of Metals Abstracts	99
19.	Instances of 0/0 ratios	102
20.	Average results for relevance categories (set 123)	105
21.	Summary of error occurrence	115
22.	Effects of searching strategies	117
23.	Partial analysis of effects of searching strategies for each month	118
24.	Precision of various strategies	119
25.	Comparison of retrieval effectiveness for different indexers	124
26.	Comparison of retrieval effectiveness according to the "difficulty" of the abstracts	174
27.	Comparative profile analysis	174

Tables (continued)

		Page
28.	Basic structures with varying numbers of parameters	175
29.	Number of forms of each type of structure	175
30.	Number of occurrences of each type of structure	176
31.	Analysis of profiles by mode of performance	181
32.	Analysis of profiles by component concepts	183
33.	Monthly profile modifications	188
34.	Frequency analysis of profile modifications	189
35.	Average number of items selected from issues	193
36.	Frequency distribution of number of items selected	194
37.	Range of number of items selected per month	195
38.	Analysis of selection change	199
39.	Overall user consistency ratios	200
40.	Frequency distribution of user consistency ratios	201
41.	Average consistency ratios	202

ABSTRACT

Research was conducted into Farradane's system of relational indexing (used to represent the explicit relationship between concepts and thereby express meaning) in order to ascertain whether or not it provided an appropriate indexing language in a selective dissemination of information (SDI) system. Five sections of Metals Abstracts, relating to the properties of metals, were adopted as the data base for an experimental system which was operated for six months. The abstracts were indexed by relational indexing techniques and a card index established. Forty-three volunteer users whose interests lay in this subject field provided interest profiles against which the data base was searched and the users also assessed the abstracts for relevance. The system was evaluated by recall - precision and recall - fallout diagrams, with Q and (M x P) being used as overall performance measures. Average system performance gave ranges of 0.69 - 0.82 for recall, 0.72 - 0.87 for precision and 0.004 - 0.003 for fallout. Various browsing techniques (generic browsing, conceptual browsing, condensations and the proposed related browsing) were used and evaluated. Relational indexing was successfully applied to more detailed indexing in a different subject area to those of previous attempts; two new uses of one relation and more complex spatial displays were adopted. Various aspects of system performance were examined, including the recall - fallout graph, and a recall - generality relationship [G(1 - M) = k] was found. The distribution of results was multi-modal, an explanation for this being sought in relation to profile structure and content. A preliminary equation for determining the number of questions for adequate testing is proposed and an error analysis was performed. The more general areas of testing methodology and user needs are discussed in relation to the SDI situation. Further research embodying methodological and human behavioural aspects is considered justifiable.

"The full meaning of a word does not appear until it is placed in its context . . . . And even then the "meaning" will depend upon the listener, upon the speaker, upon their entire experience of the language, upon their knowledge of one another, and upon the whole situation."

Colin Cherry: On Human Communication

1. INTRODUCTION

The quotation from Colin Cherry's book, "On Human Communication" [12], sums up both the ambitions and limitations of the project to apply relational indexing to selective dissemination of information (SDI). One of the objectives of relational indexing is to place a word or concept in its context by giving explicit expression to its relationships with other words or concepts. But, as Cherry points out, this is not sufficient to express the full meaning of the word, and for the successful communication and interpretation of meaning, many other factors which cannot be incorporated into the methodology of relational indexing, or any other indexing system, are present. Consideration of the SDI communication chain (shown in Fig. 1) shows the number of stages in which an interpretation of meaning occurs and it is not surprising that without including all factors, the communication process from the originator to the ultimate user will be less than perfect. Thus the objective of this project was to ascertain the effect of relational indexing's attempts to express meaning by explicit contextual relationships on the communication process, in this case specifically the SDI process.

The methodology of relational indexing involves the analysis of the subject content of a publication, its abstract or title. Not only are the concepts listed, as in many other indexing methods, but relations are used to link the concepts in such a way that the meaning is more clearly defined. The structured display

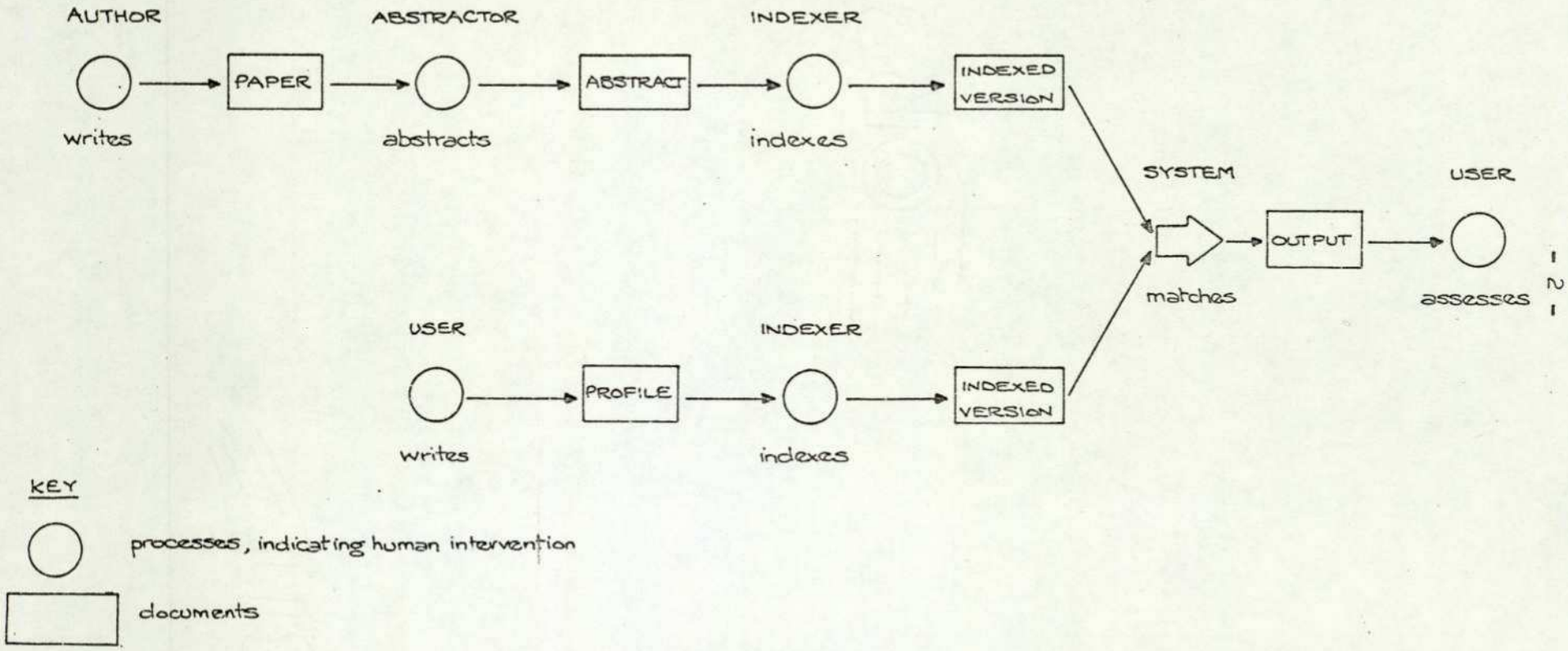


FIGURE 1 THE SDI COMMUNICATION CHAIN

is usually two-dimensional, but may be more complex. Other attempts to use explicit relations have been few and the most notable are those of Gardin [40], Austin [4, 5], Perreault [73] and Selye [91].

The system of relational indexing with which this project is concerned was devised by Farradane during the period from approximately 1950 to 1963 [26-29]. Farradane realised the need to express the meaning of a document (or other item of information) more precisely than the keywords or enumerative classification methods then in practice were capable of doing. He believed that the semantic element in an information storage and retrieval (ISR) system was of paramount importance and that the only satisfactory method of handling this element was to utilize the psychology of thinking. (In contrast Gardin's and Austin's systems were based on linguistics and general systems theory respectively and Perreault's and Selye's approaches were empirical.) As the communication stages depicted in Fig. 1 are essentially based on thinking processes, it seems eminently reasonable to use such processes as a basis. Farradane's early work was based on that of Piaget, Vinacke, Isaacs, etc, but it later received support from the cubic "Structure of Intellect" model proposed by Guilford [47, 48] (see Fig. 2). This model also supported Farradane's ideas on the structure of a general classification scheme. Later publications by Guilford and his colleagues [including 49 and 67] have developed the model and its applications (in the educational field), from which further associations of this work to that on relational indexing can be seen.

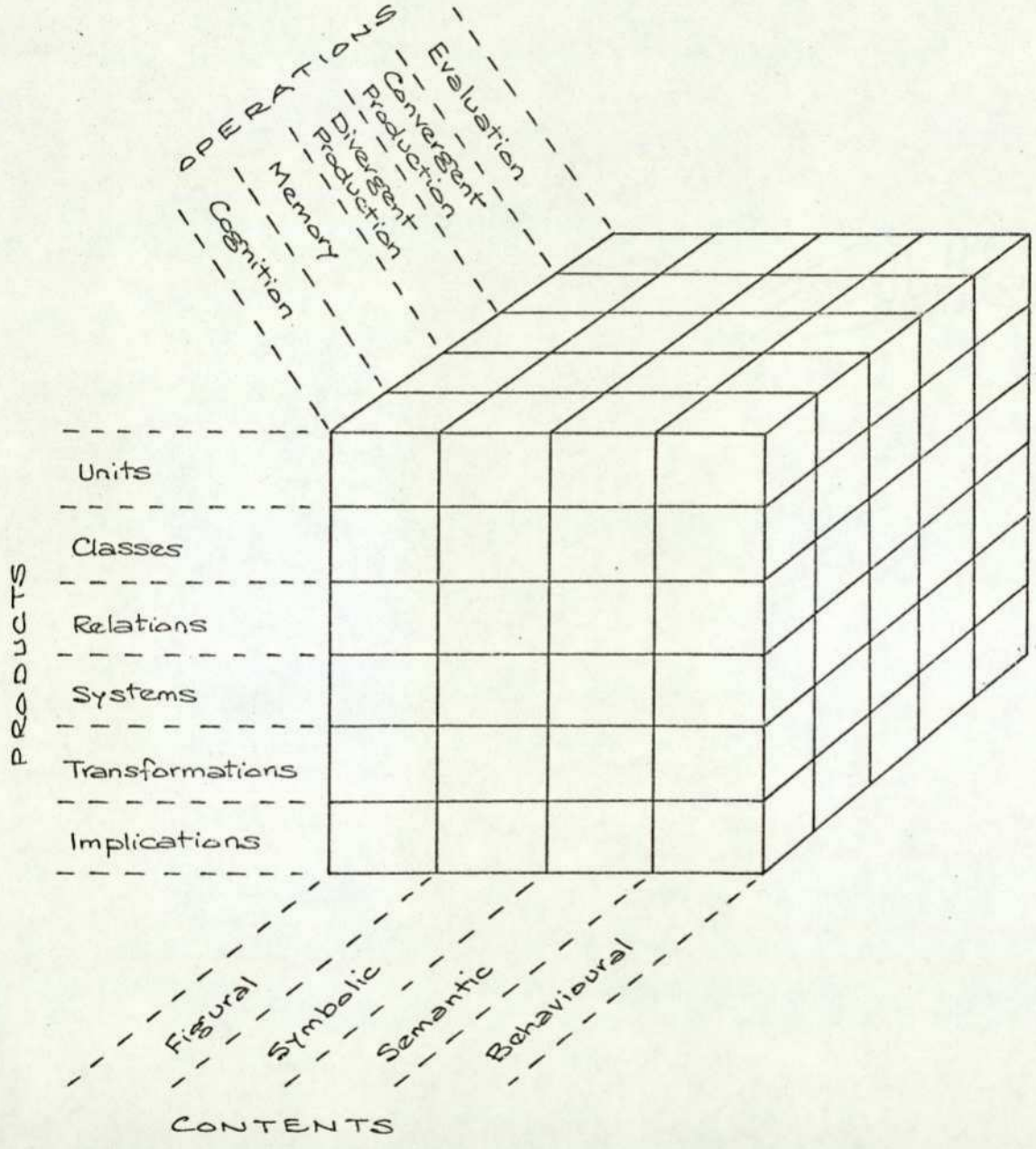


FIGURE 2

GUILFORD'S CUBIC "STRUCTURE OF INTELLECT" MODEL

In 1967 Farradane [31] wrote a more complete and explicit account of both his general classification scheme and the relational operators, with examples of their meanings and interpretations. It is considered desirable to repeat here the diagram depicting the nine relations (Fig. 3) and to indicate the range of their meanings.

Increasing Association (mental time)——→

Increasing clarity of distinction ↓	Cognition (Awareness)	Memory (Temporary memory)	Evaluation (Fixed memory)
Recognition (Concurrent)	Concurrence /θ	Self-activity /*	Association /;
Convergent thinking (Not distinct)	Equivalence /=	Dimensional /+	Appurtenance /(
Divergent thinking (Distinct)	Distinctness /)	Reaction /-	Functional dependence /:

Fig. 3 Table of relational operators

The meanings of the relational operators can be exemplified as follows:

Concurrence /θ

This implies the concurrence or juxtaposition of two concepts. It is used for the relationship of bibliographic form, e.g. physics /θ textbook, machine /θ patent. Again it is used for two things which exist together, or A in the presence of B, as B /θ A. It is a relation that has been little used in physical science documents and so far no different types of this relation have been established.

Self-activity /*

(a) The intransitive verb situation, e.g. "alloy /* crystallizing" for the natural activity of crystallizing as distinct from the crystallizing process.

(b) The dative case, where something is going "to" somewhere or something. e.g. "lending a book to a student" would be analysed as: student /* book /- lending.

Association /;

(a) The unspecified association, which is present between two concepts, whose clarity of distinction is small. Farradane's example of "prison /; disgrace" adequately illustrates this type of relation. A further example might be "rate /; shape" where the shape of an object depends on the rate of some process or activity.

(b) Indirect, calculated or abstract properties, for example, the "hardness of iron" as: iron /; hardness, the "toxicity of a chemical" as: chemical /; toxicity, the "beauty of a carving"

as: carving /; beauty.

(c) The agent of a process, for example, "heating with a flame" as: heating /; flame, "studying with a microscope" as: studying /; microscope.

(d) A past or potential process, for example, "plastic /; moulding" for moulded plastic, "book /; reading" for a book that has been or may be read, "new house /; building" for a house that it is intended to build.

Equivalence /=

(a) Trade, brand or equivalent names, e.g. alloy /= Permalloy, vacuum cleaner /= Hoover.

(b) Use as something, for example, "grain boundaries acting as sources and sinks" would be: grain boundaries /= sources and sinks.

Dimensional /+

(a) Position in time and space, e.g. research /+ The City University, accident /+ 5 pm.

(b) Temporary (variable) properties, e.g. cooling /+ rate, impurities /+ number.

(c) Temporary state, e.g. metal /+ crystal for "crystalline metal", NaOH /+ solution for "a solution of sodium hydroxide".

Appurtenance /(

(a) The genus - species and part - whole relationships. In fact in use the genus - species relationship is not always expressed as such, particularly in those cases where the relationship is not pure, but the genus is in reality an artificial class term. Thus an alloy whose magnetic properties, including magnetic susceptibility and coercive force, were measured would be represented as:

$$\text{alloy} /\left(\begin{array}{l}\text{magnetic properties}\\ \text{magnetic susceptibility}\\ \text{coercive force}\end{array}\right) /\text{- measuring.}$$

however, the appurtenance relation is used for part - whole relationships, e.g. alloy /($\{^{Cu}_{Ni}$ for Cu-Ni alloys, atom /(electron.

(b) Physical (intrinsic) properties, e.g. copper /(electrical conductivity, water /(density.

Distinctness /)

(a) For imitations, e.g. surface /) replica for "a replica of the surface", pearl /) imitation pearl.

(b) For substitutions, e.g. carbonate /) phosphate for "phosphate substituted for carbonate".

(c) The awareness of differences or distinctions, e.g. free molecular flow /) flow for "flow which is different from free molecular flow", needle-like /) equiaxed for "equiaxed as distinct from needle-like".

Reaction /-

Varying applications of this relation have yet to be established. It is used for activity or process concepts occurring at the present time or for one thing acting on another thing or process. Apart from distinguishing between things and processes or different types of processes, for which no need has been felt, the relationship seems to be the same in each case. Examples are: metal /- heating, density /- measuring, iron /- acid.

Functional dependence (or causation) /:

This is used for A arising out of B or B causing A, such that B /: A. Examples of this relationship are: iron ore /: iron, tree /: leaf (fallen) for "a leaf that has fallen from a tree", bacteria /: disease. It is a relation that has not been excessively used in physical science documents and, like reaction, other uses of this operator have yet to be discovered.

The relational indexing system has now been used in a number of projects. Several small researches have been conducted by M.Sc. students of information science at The City University, investigating such aspects as condensations [84, 101], expansions [55], inference steps [11], conceptual browsing [78] and the use of relational indexing for indexing German language documents [64]. Relational indexing was also used as one, not fully tested, part of the index language evaluation at Aberystwyth by Keen and Digger [56, 57]. However, up to now, the largest evaluation so far conducted was that by Farradane, Datta, Poulton and Hills [23, 32, 36]. In the latter

research, a relationally indexed data base of sugar technology abstracts was tested as a retrospective search system. Several parameters were used to evaluate the performance of the system and the results appeared very promising [33, 37]. The performance was considerably better than is frequently obtained from ISR systems. It was the potential nature of these results that led to the present research project being proposed. The initial question raised concerned the performance of relational indexing in the SDI situation which is known to be somewhat different from the retrospective search system. Additionally, it was anticipated that a further test of relational indexing with a different subject data base would help to establish or otherwise the long term applicability of relational indexing to a variety of subject matters. The third objective of the research was to consider ISR systems and to further the knowledge of the principles of their operation and evaluation if possible.

At this point it should be noted that as the number of SDI systems that are, or have been, operational or have been set up for testing purposes is relatively high, it is not proposed to review them here as they have only limited relevance to a system based on relational indexing. Where appropriate, mention is made of the more significant of these systems throughout the text.

2. METHODOLOGY

The purpose of an SDI system is to select periodically current information from a data base to meet the information needs of a user. The needs of the user are expressed in, what is usually termed, his "profile" of interests and the user may be one individual or a group of people with similar interests, in which case the needs are expressed in a "group profile". In respect of the research project being reported here, all the users were individuals.

The SDI situation is somewhat different from the retrospective search situation. In the latter case, the user poses a question and he usually requires the total data base to be searched in order to select items of information which are relevant to his information need. Very often he is interested in obtaining an item of information which will answer his particular question and is not so concerned that all items having a degree of relevance to the question posed should be selected, although obviously he has no desire to receive irrelevant material in the process. Using Fairthorne's phraseology [25], this represents the "only but not all" (OBNA) situation. For SDI, a search is made of current, new information that has been added to the data base since the last search was made. The data base for each search is relatively smaller than that used for a retrospective search. The user usually wants all the information that is relevant to his information needs, and he is willing to accept a number of "fringe" papers, which although not directly relevant, may stimulate new thinking. This represents the "all but not only" (ABNO) situation.

Discussion of which is the "more difficult" situation to test seems irrelevant. They both have their associated problems, some of which are similar, some of which are different. Another area of debate that arises is how close the testing situation is to real life, and therefore whether or not the results are valid if the parameters are not closely related. From this point of view the results obtained from a system set up specifically for simulation purposes may invariably be less convincing to others than those obtained in an operational environment. Yet testing an operational system causes a disturbance which may itself render the results less accurate than would otherwise be hoped.

2.1 Testing methodology

Current thinking on testing methodology depends largely on the ideas initiated by Documentation Inc., Armed Services Technical Information Agency (ASTIA) and Cleverdon in the early 1950's [15], namely posing questions to a system, producing an output in relation to the question and evaluating that output. The methodology was more developed by the time of the second Cranfield project [18] but, although the details of the procedure may vary from test to test, the basis has remained the same in the two decades that people have been testing information retrieval systems. Generally speaking, these techniques seem a reasonable, logical way in which to tackle the problem of evaluating an ISR system, but closer examination reveals weaknesses in the methods.

One of the main problems relating to the methods, when compared with "scientific" methods, is the large number of variables which it is virtually impossible to control. The major groups of these uncontrolled variables, associated with the assessment of relevance, have been enumerated in several publications and are discussed in greater detail in Section 3.3.8. They include all the human factors associated with judgement and decision making processes. There are also a number of human factors associated with the system itself. In a system using a controlled language, the human variables are basically the same as those of the user, and are just as unmanageable. Even with systems using reputedly natural language, there are a number of human intervention steps that can incorporate such variables.

Therefore information retrieval systems as they exist at present embody a large number of uncontrolled variables and any results obtained in a testing situation probably incorporate them to such an extent that they may mask any true performance factors associated with the systems.

But the fact remains that no satisfactory and acceptable alternatives had been proposed at the beginning of this project in 1969 and are still not available at the time of writing. Thus the methodology outlined in the succeeding sections relies on the usually adopted approach of Cleverdon and others.

2.2 Data base

In conducting this research, the material used to form the data base was five sections of Metals Abstracts. The sections chosen were 1.2, Crystal Properties, 1.4, Structural Hardening, 3.2, Physical Properties, 3.3, Electrical and Magnetic Phenomena and 3.4, Chemical and Electrochemical Properties. This particular abstracts journal and the aforementioned sections were selected in order to combine a fairly coherent body of literature, in which a reasonable number of people were likely to be interested, together with some proximity to the author's original discipline of physics. Based on an analysis of back issues of the abstracts journal, the number of abstracts in these sections seemed likely to give a data base of approximately the required size. It was considered that about 1500 abstracts would be suitable for testing, and a number that would give a fairly large data base, without being beyond the capabilities of one researcher to handle, in conjunction with acceptable time scales and subsidiary activities. It was also felt that a six month period should generate sufficient results to test various factors in relation to data base size, would enable profiles to undergo minor modifications and reach a stable point within this time period, would not unduly prolong the length of the project, and would not be excessively demanding on the goodwill of the volunteer users. Unfortunately there is no statistical data available to inform prospective system evaluators of the size of data base, time scale, etc. they should employ in order to obtain viable results. In fact, as it subsequently emanated, the six month time period may have been too short in some

respects.

Allied to the above decisions, the most appropriate time period for the experimental part of the project appeared to be July - December 1970. The number of abstracts published per month seemed to be higher in the summer than in the winter and consequently it was considered that this half of the year was more likely to provide a general cross-section result. Based on the previous rate of publication these six issues should have provided between 1500 and 1800 abstracts. The month of June 1970 was also used, not however as part of the main project, but as a trial run with some of the users, in order to become acclimatised with the procedures and check that they were working satisfactorily. Unfortunately the Metals Abstracts publishing programme became a little erratic, with the July 1970 issue being considerably larger than usual and additionally being published late. This upset what was originally a very tight time schedule (of running the project as a truly current system), a delay from which the project never quite fully recovered. Additionally both the August and November issues were larger than had been anticipated by comparison with previous years. The actual number of abstracts published in each section over this period is given in Table 1.

It can be seen from Table 1 that the number of abstracts involved in the six month period of the project was 2825, and in total, including the month of June, some 3202. It should also be noted that Metals Abstracts records publication of papers taken

TABLE 1

Number of abstracts published in five sections of Metals Abstracts in June to December 1970

Month (1970)	Sections					Total
	1.2	1.4	3.2	3.3	3.4	
June	91	21	48	146	71	377
July	181	57	94	262	129	723
August	152	33	67	174	22	448
September	77	26	39	123	24	289
October	64	22	52	124	36	298
November	160	44	87	191	105	587
December	140	37	80	126	97	480
Total (excl. June)	774	219	419	1000	413	2825
Total (incl. June)	865	240	467	1146	484	3202

from the English translated version of foreign journals. If abstracts of these papers have been previously included in Metals Abstracts from the original foreign publication, the reference is given again with an entry in the form "See Met. A., ...", referring to the actual abstract. These entries were not included in the data base nor in the figures given in Table 1. They have been completely disregarded throughout the project as such abstracts

were not immediately available to either the evaluator or the users.

On receipt of the issues the appropriate abstracts were indexed using relational indexing, and the indexed version was a two dimensional display of concepts and relations. These indexed formats were then typed on to spirit duplicating masters and copies were run off on to cards. The terms considered useful for retrieval were indicated and a heading given to each card, which consisted of the particular term (underlined) and the term and relation immediately preceding it. In cases when two or more terms preceded the particular term, a card was made for each combination. These cards were then filed in alphabetical order of the underlined terms according to filing rules devised. Typical card index entries for one abstract are shown in Fig. 4. During the course of the research some 32,000 cards were raised. (Examples of indexed abstracts are given in Appendix III.)

2.3 "Users"

Once a suitable data base had been decided upon, the problem arose of finding an appropriate number of volunteers to act as users for the duration of the project. Again the published literature surveyed provided no guidelines as to the appropriate number of users for an evaluation of this type, and initially it was considered that about 100 users, combined with the anticipated 1500 abstracts, would provide relative conditions to those of other ISR system testing. Initially 30 invitations to participate were sent to research associations, industrial organizations and

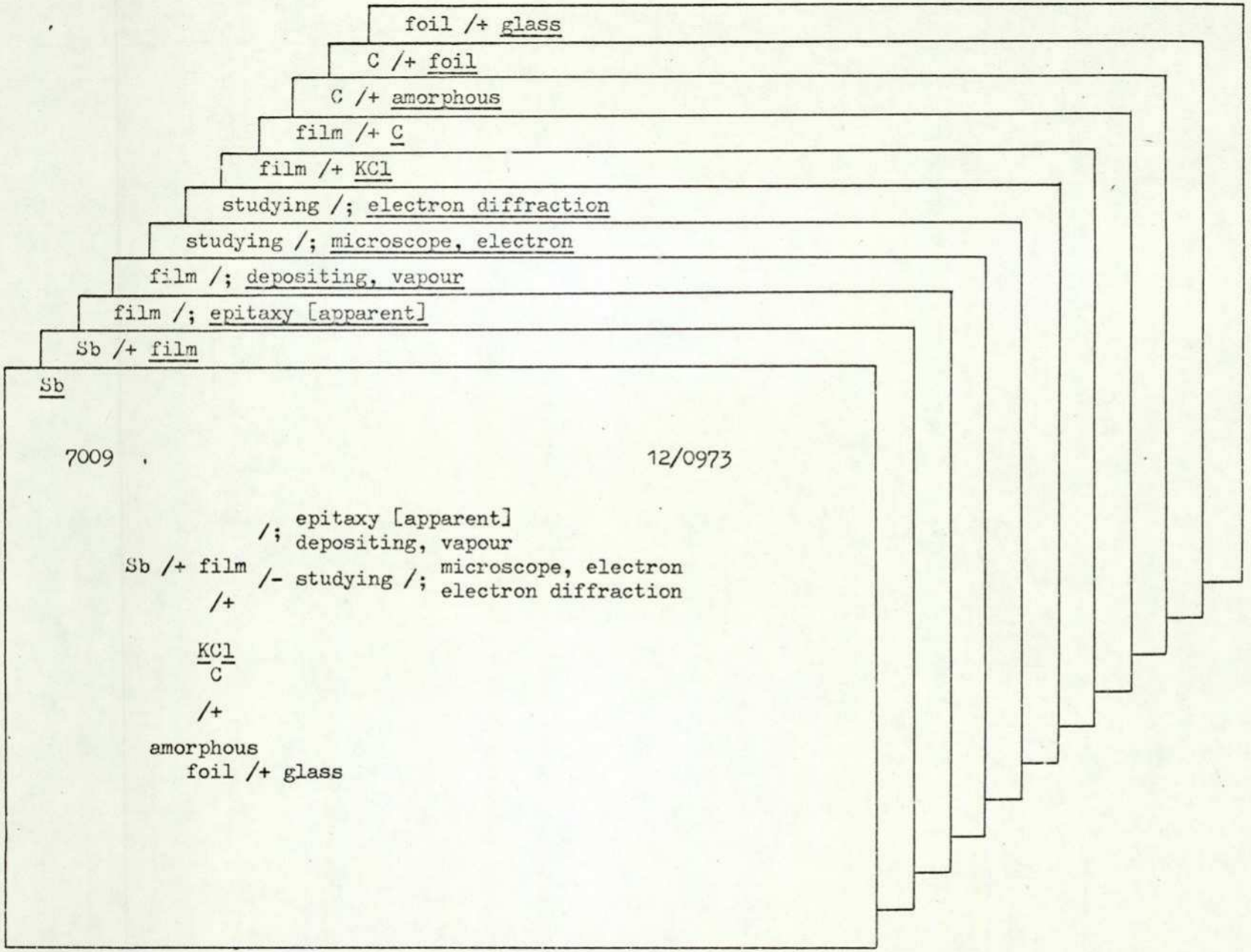


Fig. 4 Typical card index entries for one abstract

university departments which had an interest in the properties of metals and were located within a 50 mile radius of London. These letters resulted in 19 volunteers. The second approach was directed to appropriate physics and metallurgy departments of universities and colleges of technology and a further 76 letters produced 23 more volunteers from these sources. In the third approach 56 letters were sent to research associations, industrial organizations, government establishments, etc., throughout the country, resulting in another 29 volunteers. Having exhausted the more obvious sources in this country, it was decided that the 71 users recruited would be adequate to embark on the next stage of the project. Later experience proved that this number was more than enough for one person to monitor. However, from the viewpoint of endeavouring to obtain co-operation for this kind of project, it is worth noting that a total of 162 letters resulted in 71 users, or an overall response of approximately 44.5%, although it must be borne in mind that some organizations did in fact provide more than one user.

As is perhaps inevitable in this type of research, many of the users dropped out during the ensuing months, some for quite legitimate reasons, for example, going abroad, changing jobs, too many commitments, etc., but others apparently through lack of interest. Of the 43 users who completed the project, the distribution according to type or organization is as follows:

Academic institutions	22 users
Research associations, institutes, government laboratories, etc.	9 users
Industrial organizations	12 users
Total	43 users

Naturally one would have wished for a more even distribution across the three groups, but it is understandable that university and polytechnic staff are somewhat more readily disposed to participate in this type of project, particularly if engaged in academic research themselves. However, it is felt that this distribution should not have adversely affected the results.

The letters inviting participation were dispatched from the end of February to the end of April 1970 and the last volunteers offered to participate in June 1970. The collection of profiles commenced almost as soon as a user volunteered.

2.4 Profile collection

The collection and recording of profiles is an area of SDI system testing which did not intially receive the degree of attention it is subsequently warranting. Obviously it is the first step in the sequence of information retrieval if the latter is thought of as a serial process and, because it necessitates human involvement, is liable to human errors. It has long been realised

that a user, when framing an information seeking question, rarely poses the question to which he actually wants an answer. Although the profile in an SDI system is analagous, one often labours under the impression that, because an SDI system aims to provide a service to the user, the user will co-operate by providing an indication of his real information need. Again because it is a service, many people considered that compiling the information profile from a user's statement of his interests was part of such a service and should be undertaken by the system organiser. Recent work on MEDLARS [61, 62] has indicated that this may not be the best way of obtaining a good profile. The increasing tendency for users to have to subscribe to commercial systems whose services they have previously been receiving free while the system was in an experimental or developmental stage, is perhaps another reason for a growing interest in accurate profile construction, as faulty construction frequently results in both actual and opportunity losses in money and time.

Relatively little has appeared in the literature on profile collection. Perhaps intuitively one feels that the best way of obtaining most accurately a representation of the user's information need is to talk to the user, or to have some form of user interaction which is as close to the heart of the system as possible. This, of course, is analogous to the situation where a user asks a question of the information department, when ideally the person who is going to conduct the search discusses the problem with the user in as much detail as possible in order to isolate

the real heart of the information requirement. At the time of profile collection for this project a similar judgement was made and this was one of the reasons why, in the first instance, an attempt was made to obtain users from within a 50 mile radius of London, in order that interest statements could be obtained personally. In actual practice an adequate number of users were not found in this area and so the search had to be extended further afield, thus rendering such an objective impracticable. Nevertheless it was still considered that a discussion with each user would give greater insight into his information need, and, where possible, personal discussions were held. Where this was not possible, the users were requested to send a detailed statement of their interests, including definitions of any terms that they considered would not be readily understood by the researcher, together with the numbers of some previous abstracts from Metals Abstracts which were relevant to their interests. This was essentially the same methodology as that used by INSPEC in the gathering of their profiles for their experimental system [13]. If the statements of interests were not completely clear, a letter was written to the user asking for further clarification before the profile construction was completed.

2.5 Profile construction

The method of profile construction is largely determined by the mechanics of the selection process. The two most frequently used methods for both controlled and uncontrolled

vocabularies are:

(a) Boolean logic, by which the terms in a prepared list are linked. The terms in each parameter are linked by OR logic and the parameters themselves are linked by AND or NOT logic. This method is used by UKCIS [97], although their actual methodology is rather more sophisticated, embodying WITH and IGNORE logic, term truncation and possibly weighting. INSPEC [1, 13] also use Boolean logic. The IBM Technical Information Retrieval Center (ITIRC) system uses a form of Boolean logic which incorporates ADJACENT and WITH and builds up sub-concepts into main concepts [70].

(b) Weighting, by which each term is assigned a "weight". In matching, the value of the document is the sum of the weights of the terms present in the document and, if this exceeds a "threshold value" set in the profile, the document is selected. The "weighting" method is generally considered preferable to Boolean logic, but some care has to be exercised in assigning weights and the threshold value. An example of this method is exhibited in the ICI ASSASSIN system [19].

The profile construction methods used in various systems may vary slightly from the basic principles outlined above. For example, Anthony, Cheney and Whelan [3] describe the Culham Laboratory system in which the terms are arranged in a matrix format according to the user's requirements. Kerr-Waller's [58]

electronics SDI system at Standard Telecommunications Laboratories used terms grouped into three groups according to their importance and, in matching, different strategies provide for the different percentages of occurring terms according to the group in which they are placed.

Some systems are not based on either controlled or uncontrolled vocabulary. An example of a system based on a classification, in this case the Universal Decimal Classification, is the British Steel Corporation's system, described by McCash and Carmichael [66]. The profiles in this system are constructed from UDC numbers. The SDI system at the National Cancer Institute uses HICLASS, a hierarchical classification developed by Schneider [90]. The US Public Health Service, Office of Pesticides, also uses classes and subclasses, rather than single concepts, for profile construction [100].

UKCIS have observed that about 35% of interest statements were defective in some way and Barker, Veal and Wyatt [8] distinguished four types of terms, namely specific, non-specific but useful, redundant, and detrimental terms. They experimented with automatically generated lists of terms to improve profiles and, although the work was done on single terms, whereas most profiles have pairs of terms, they concluded that the work was sufficiently worthwhile to extend it to pairs of terms. Heaps and Ko [52] at the University of Alberta are also working on automatic question formulation.

In the present project the profiles had to be constructed as two-dimensional displays of concepts interlinked by relational operators. This is necessary so that a complete structure, incorporating both concepts and relations, can be matched against the items in the data base. The system incorporates certain elements of Boolean logic in so far as a square bracket represents an OR situation and a curly bracket indicates an AND situation in which all the linked concepts must be present. It is implicit within the relations themselves that an AND situation is represented. One aspect of relational indexing is that no satisfactory means of weighting search terms has yet been found and thus that an output ranked on weights or threshold values has not been obtained. Earlier efforts [32, 36] at ranking output, by the number of additional concepts and assigning values to the different relations, have been attempted, but these were not successful.

2.6 Matching

As described in Section 2.2, a series of cards were produced for each indexed abstract and these were filed alphabetically, according to the entry point concepts, in index card cabinets, with one set for each month's issue of Metals Abstracts.

The matching operation consisted basically of selecting a concept from the profile, searching under this concept in the file and comparing the complete structure with that of the profile for items containing the appropriate concept. To minimize the work

involved the concept chosen from the profile was that considered likely to produce the smallest subset of the file to search through, i.e. the most "specific" term in the sense of "specificity" used by Sparck Jones [93]. Care was exercised throughout the procedure that any profile relationships effectively representing the Boolean OR logic were fully covered by the search.

However, within this basic framework there are a number of different strategies that can be applied to the matching process. These strategies are specifically associated with relational indexing although in some cases there are parallels with keyword matching.

The basic strategy is a "complete match" between the profile and the item in the data base, i.e. both concepts and relations correspond.

The second strategy has been termed "match plus" and represents the situation in which the item in the data base not only completely matches the profile but also contains additional concepts and relations. Theoretically such selected items should represent a narrower aspect of the profile than that produced by a complete match. However, in this project the match plus strategy was widely used and signified the difference in exhaustivity of indexing between the abstracts and the profiles.

The difference in depth of indexing between abstracts and profiles was counteracted by "generic browsing", a technique

by which the basic structure of the profile was compared with the data base, but one or more of the concepts in the profile was replaced by more generic or more specific concepts. This technique was also frequently adopted because users tended to include in their profiles such concepts as "magnetic properties" which in reality implied an interest in papers on magnetic susceptibility, coercive force, magnetic moment, magnetization, etc.

"Conceptual browsing" was also employed and is an effective technique for retrieving relevant items, although the details of its operation are not fully explored at the present time. This technique is discussed in further detail in Section 3.1.8.

A fifth strategy used was "condensations" or "logical jumps", and the details of this method are more fully developed [38]. The basis of the method is that an item is selected if its structure (or that of the profile) will condense, according to a predetermined set of rules, to become a complete match with the other. Again the difference in exhaustivity of indexing between the abstracts and the profiles meant that it was the abstracts rather than the profiles which were condensed.

A new strategy was also employed. This was "related browsing", which is essentially the same as generic browsing, the main difference being that instead of replacing a concept by a more generic or more specific term of the same type, it is replaced by a term which is obviously related, but of a different concept type,

e.g. "crystal" might be replaced by "crystallizing". Related browsing is discussed in greater depth in Section 3.1.10.

All of the strategies described above were used to select items against the various profiles. The numbers of selected abstracts were listed but were not sent to the users.

In order to obtain a value for the number of relevant abstracts in an issue of Metals Abstracts, the users were asked to scan the appropriate sections of each issue and note down, on a form (see Appendix I) the numbers of those abstracts of interest to them. The form was divided into two parts, namely "Abstracts of definite interest" and "Abstracts of possible interest" and users were requested to enter the abstract numbers in one of these two columns. The forms were sent out monthly and, when completed, were returned. There were inevitable delays in receiving the completed forms because of the availability of time and Metals Abstracts to the users. However, because of the work load in indexing and preparing the cards, the delay was not of major importance.

2.7 User assessment

The lists of system-selected and user-selected abstracts were compared in order to give a value for relevant items retrieved. The non-matching documents retrieved, both by the system and the user, were re-examined. In some cases the reason for the discrepancy was immediately apparent. For those

abstracts where the reason was not obvious, a list of the discrepancies was sent to the user who was asked to comment thereon. In some cases the user revised his selection; in others he explained, with varying degrees of detail, why the abstract concerned was relevant or non-relevant to his interests.

The revision of the user's selection raises a number of points. If he eliminated abstracts from his original selection, he often commented that he was unable to give a justifiable reason for the original choice. If, on the other hand, he agreed that non-matching items selected by the system were relevant, it implied that his own scanning was not perhaps as complete as it might have been. Doubt is then cast on the figure for total relevant documents in the system - perhaps there were others which were missed by both the user and the system. On the other hand, in a true SDI environment the user would have been sent a list of notifications according to the system selection and hence it would have been this set that he assessed.

The attempt to divide the users' selections into "definite" and "possible" interest was fraught with difficulties. Some users will say that everything is of definite interest, while others will say that the vast majority are of possible interest and would rarely commit themselves further without seeing the full text of the paper. The approach adopted in this project did not seem to be any more or less viable than other attempts to provide a relevance dichotomy.

The methodology adopted to obtain a value for total number of relevant items was viable in a project of this nature. One reason was that some of the users were already scanning Metals Abstracts regularly and hence the examination of these five sections involved them only in the extra work of entering the numbers of relevant abstracts on the form. Other users were not doing this scanning and in fact some of them did not have Metals Abstracts available to them, so that photocopies of the necessary sections were supplied. Also, although the total data base for the project consisted of about 3000 abstracts, still small compared with very many systems, the task of critically scanning the total data base would have been rather large for a volunteer user to undertake. However, as the data base was studied in six or seven portions, each of about 400 abstracts, it was not too great a task to ask one person to undertake. One problem with the project, using this technique, was that one had nothing to offer the user in return. This may have been one of the factors affecting the apparently poor response to the letter of invitation to participate in the project. In many of the other SDI evaluations, the organisers were able to offer an SDI output to users in return for some form of evaluation on their part.

2.8 Evaluation measures

Measures for the testing of ISR systems is an area of information science in which there appears to be more discussion than constructive results. Various measures have been proposed at

various times by different investigators, but the only two that have really gained recognition, albeit not acceptance in all quarters, are recall and precision.

The suggested quantities can be roughly grouped according to the aspect of the system they endeavour to assess, for example, coverage, currency, search time, cost, user effort involved, retrieval performance. While it is recognised that all these aspects are important bases on which to assess ISR systems, particularly operational services, many of them were not considered to be helpful in evaluating relational indexing for SDI systems. In this project the emphasis was placed on retrieval performance.

The basis of many of the performance evaluation measures is the 2 x 2 contingency table, borrowed from statistics and modified to the information retrieval situation, as follows:

Figure 5

2 x 2 Contingency Table

		SYSTEM		
		Retrieved	Not retrieved	Total
USER	Relevant	R	C - R	C
	Non-relevant	L - R	(N - C) - (L - R)	N - C
	Total	L	N - L	N

(Some authors replace R. L - R, C - R and (N - C) - (L - R) by a, b, c and d respectively, or other symbols, but in this discussion the given notation is used and other symbols are transcribed into

this notation.)

Some of the parameters that have been suggested, including those derived from this table, are:

(i) Recall, $M = \frac{R}{C}$ [14]

(ii) Precision, $P = \frac{R}{L}$ [14]

(iii) Fallout, $F = \frac{L - R}{N - C}$ [18]

(iv) Generality, $G = \frac{C}{N}$ or $1000\ \frac{C}{N}$ [18]

(v) Sensitivity = recall [75]

(vi) Specificity = $\frac{N - C - L + R}{N - C}$ = 1 - fallout [75]

(vii) Normalised recall = $1 - \frac{\sum_{i=1}^{C} r_i - \sum_{i=1}^{C} i}{C(N - C)}$ [85]

where r_i = rank order of the ith document

(viii) Normalised precision = $1 - \frac{\sum_{i=1}^{C} \ln r_i - \sum_{i=1}^{C} \ln i}{\ln \frac{N!}{C!(N - C)!}}$ [85]

(ix) Recall + precision = M + P

(x) Effectiveness = sensitivity + specificity = R + 1 - F [75]

(xi) Effectiveness = sensitivity + specificity - 1 = R - F [46]

(xii) Q value = $\frac{R(N - C - L + R) - (C - R)(L - R)}{R(N - C - L + R) + (C - R)(L - R)}$

$= \frac{M - F}{M + F - 2MF}$ [36]

(xiii) Measure of merit = $R - (L - R) - (C - R) +$

$[(N - C) - (L - R)]$ [98]

$= 4R + N - 2L - 2C$

(xiv) Effectiveness = $100\ (\frac{R}{C} - (1 - \frac{R}{L}))$ Borko, quoted by Cleverdon and Keen [18]

(xv) $F = \dfrac{100\ \frac{R}{S}}{\frac{(L - R)}{S} + \frac{(C - R)}{S} + 1}$ Vickery, quoted by Cleverdon and Keen [18]

where S (normalisation factor) = $R + (L - R) + (C - R)$

(xvi) $\lambda = \dfrac{4R}{L + C} - 1$ [83]

(xvii) $I = \dfrac{R \log NR}{CL} + \dfrac{(N + R - L - C) \log N(N + R - L - C)}{(N - C)(N - L)} +$

$\dfrac{(C - R) \log N(C - R)}{C(N - L)} + \dfrac{(L - R) \log N(L - R)}{L(N - C)}$ [68]

(logs to the base 2)

(xviii) Efficiency, $\eta = \dfrac{1}{\sqrt{[(\frac{1}{M} - 1)(\frac{1}{P} - 1) + 1]}}$ [81]

(xix) Focussing, $\theta = \dfrac{2MP}{(M + P)}$ [81]

(xx) $E = \dfrac{\mu_2 - \mu_1}{\frac{1}{2}(\sigma_1 + \sigma_2)}$, where μ_1 and μ_2 are the means of the density functions of relevant and irrelevant items and σ_1 and σ_2 are their standard deviations [95]

(xxi) $S = \dfrac{\mu_2 - \mu_1}{(\sigma_1^2 + \sigma_2^2)^{\frac{1}{2}}}$ [10]

(xxii) Extension ratio (of users' literature awareness)

$$= \frac{b}{a + m + 1} \quad [69]$$

where b = number of relevant documents retrieved by system and not known before the search

a = number of relevant references retrieved by the system and known before the search

m = number of relevant references not retrieved by the system but known before the search.

(xxiii) Expected search length $(\overline{esl}) = \frac{1}{N} \sum_{n=1}^{n} U_n + \frac{1}{2} V_n$ [21]

where U_n = number of documents of higher rank than document relevant to the n^{th} request

V_n = number of other (irrelevant) documents of same rank as the relevant document.

(The number of irrelevant documents the user would be expected (in the sense of statistical expectation) to have to search through before finding as many relevant documents as he needs. The average $\overline{esl}$ is the average of the individual esl of all users of the system.)

(xxiv) Product, (M x P) [99]

(xxv) Attraction and repulsion coefficient

$$= \frac{R - (C - R)(L - R)}{\min [(C - R),(L - R)] - (C - R)(L - R)} \quad [59]$$

(xxvi) Normalised "sliding" ratio measure [74]

The most frequently used representation of performance is the recall - precision graph used by Cleverdon and Keen in Cranfield II [18]. This has the advantage of representing factors that can be readily understood, but it has been criticized by several people, including Fairthorne [25], as being mathematically unsatisfactory, mainly because it does not include all parameters of the 2 x 2 table. The recall - fallout diagram is more suitable from this point of view and it is being increasingly adopted, though not to the detriment of recall - precision. Generality is another parameter that has received some consideration and Robertson [80] considers that it should be quoted as part of the experimental conditions.

Of the other measures quoted above, the majority are unsatisfactory for one reason or another. Normalised recall (vii), normalised precision (viii), expected search length (xxiii) and normalised "sliding" ratio (xxvi) are not applicable as they require a ranked output. Doubt has been expressed [25, 80] about any overall measure which does not include R, L, C and N and, on this basis, recall + precision (ix), effectiveness (xiv), F (xv), λ(xvi), efficiency (xviii), focussing (xix), product (xxiv) and attraction and repulsion coefficient (xxv) would be eliminated. Neither of the values of effectiveness, (x) and (xi), nor the measure of merit (xii), all proposed by the Case Western Reserve University, appear particularly helpful as they will be close approximations to their largest constituent, e.g. R + 1 - F $\simeq$ .1, R - F $\simeq$ R, measure of merit $\simeq$ N. Meetham's I (xvii),

Swets' E (xx) and Brookes' S (xxi) are all subject to the question: can values obtained in an investigation such as this be regarded as probability functions? Miller's extension ratio (xxii) is as much a measure of the user as of the system. Therefore in this investigation the results are depicted as recall - precision and recall - fallout diagrams, with Q values and the product (M x P) used as overall measures, and consideration given to generality.

3. RESULTS AND DISCUSSION

3.1 Relational Indexing

3.1.1 The relations

During the course of this research project certain views were formed on the relations and their use. Some of these were adopted during the project; others are put forward here for consideration in the future.

Self-activity /*

In Section 1 two uses of the self-activity relation were given, namely the intransitive verb and dative case situations. A third use was considered to be the expression of "through" in, for example, the case of "heat transfer through a metal" which would thus be "heat transfer /* metal". This type of relationship is of both a temporary memory and low clarity of distinction type and hence the self-activity relation seems appropriate, especially in view of its position above the dimensional relation in the table of nine relations.

It is also tentatively suggested that certain properties of a phenomenological nature could be represented by this relation. Such concepts as "ferromagnetic resonance" seem to be sufficiently indistinct and unevaluated in our present state of knowledge not to warrant a relation further down or to the right of the table. Also they seem to hover somewhere between being properties and

activities, but have certain characteristics of both, at least in scientific writings. It is for such reasons that the self-activity relation seems to be the most suitable for their expression.

Association /;

Of the four main uses of the association relation, it is the use for past and potential processes that seems the most debatable. There would appear to be a certain lack of clarity in representing both past and future (potential) processes by this relation, as a process that has been applied is both more distinct and more evaluated than the relationship between a concept and a process that might be applied sometime in the future. For example, the relationship between a house that has been built seems much more definite and evaluated than that between a house and building which may or may not take place. Perhaps the dimensional relation might be a more accurate representation of the past process, inasmuch as the concept to which the process has been applied is in a certain state, and the relation is similar to that of a process that is taking place now, except that it is transported back in time.

The future (potential) process provides a more difficult problem and it is thought that the concurrence relationship may in fact be more correct in this situation as the relationship is both indistinct and unevaluated. However, no difficulty has been experienced in the present project where the same operator was used for both meanings.

The expression of time relationships is difficult with relational indexing as with other indexing systems. One can distinguish between the present and the past, but not between different events that happened at different times in the past. This may not always matter, but there are situations in which it could be important. One example of such a situation would be the indexing of a patent of a process containing several steps; the order of which was important. Obviously one can circumnavigate such problems by such devices as numbering the steps in the order in which they occur, and equally obviously it would not be a practical proposition to include different relations for each past process that occurred, as the semantic relationship is essentially the same in each case. Perhaps the most correct way of using relational indexing to express such relationships would be as follows:

```
                                                    or
                                                    object /- process 1
                                                      /:
                                                    product 1 /- process 2
                                                      /:
                                                    product 2 /- process 3
                                                      /:
                                                    product 3
object /; { process 1
            /+
            before
            /+
            process 2
            /+
            before
            /+
            process 3
```

although both of these techniques are rather laborious.

3.1.2 Number of relations associated with one concept

Earlier in the history of relational indexing it was though that there was an empirical maximum of four relations

associated with any one concept [30]. These relations should be evenly distributed in two dimensions, i.e. with the concept subordinate to two relations and with two relations subordinate to the concept, as shown below:

$$/_1$$

$$/_2 \ A \ /_3$$

$$/_4$$

where A is the concept and $/_1$, $/_2$, $/_3$ and $/_4$ represent different relations.

In this research there were a number of instances when a concept had four relations associated with it, which were not distributed in the manner shown above. The four relations may all be subordinate to the concept, thereby negating any ideas that human thinking, as depicted by relational indexing, is restricted to two dimensional representation.

A survey was made of the examples in which more than four relations were associated with any one concept. In the six months from July to December 1970 one hundred examples were found of such cases, which is a reasonably high number although of course a small percentage of the total number of abstracts indexed. Of these one hundred examples, 88 involved five relations, 9 involved six relations and three involved seven relations. The majority of the cases (63) had all the relations subordinate to the concept, whilst in 37 of the examples the concept was subordinate to one or more of the relations. In 28 of the latter

cases there was one relation before the concept, in eight cases two relations before the concept and in one case, three relations before the concept. Thus in complex analyses there would not seem to be any limit (other than nine) to the number of relations that can occur either before or after the concept in question.

In the one hundred examples, the frequency of occurrence of the various relations in both the "before" and "after" positions was studied, and the analysis is given in Table 2.

TABLE 2

Frequency of occurrence of relations

in examples with more than four relations

	No. of times the relation occurred before concept	% of examples	No. of times the relation occurred after concept	% of examples
/0	-	-	8	8
/*	3	8	70	70
/;	9	24	79	79
/=	4	11	52	52
/+	9	24	88	88
/(	16	43	82	82
/)	1	3	-	-
/-	-	-	58	58
/:	5	14	31	31

From Table 2 above, it can be seen that /(, followed

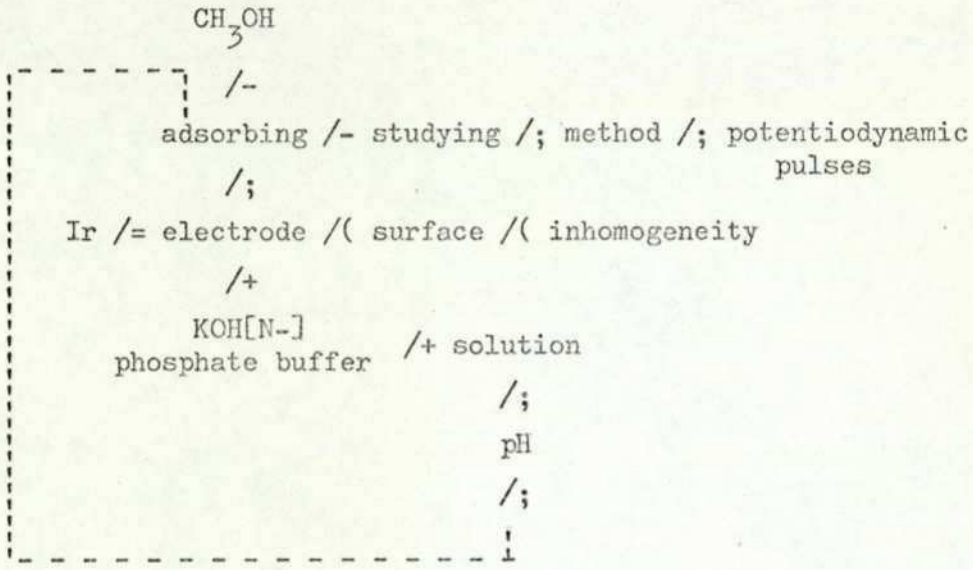
by /; and /+, are the relations which most frequently subtend the concept. Also /+, /(and /; most frequently follow the concept, each of these relations occurring in over 75% of the examples. The other relations that follow a concept in this type of structure are /*, /- and /= and less frequently /: and /0. No instances of /) were found in this set. Thus, if five relations are subordinate to a concept, these are most likely to be /+, /(, /;, /* and /- (or possibly /=). If one relation subtends the concept then this relation is most likely to be /(and the four following relations are then /+, /(, /; and /*. In terms of the subject area studied, these results are not surprising; the abstracts are mainly concerned with alloys, which may be part of something (hence the /(before the concept), which have constituents [/(], are in a temporary state, e.g. crystal [/+], have some associated property [/;], do something (self-activity) [/*] and have something done to them [/-] or are equivalent to something else [/=].

3.1.3 Three dimensional representation

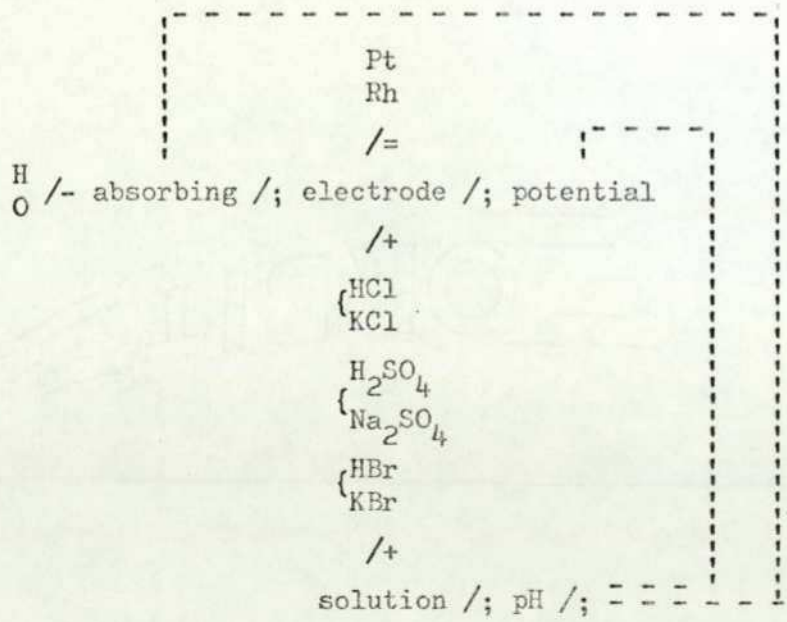
In the earliest work on relational indexing, the initial representation of a subject was linear (any complications being represented by brackets inserted in the linear notation). This method was later discontinued (although there is no inherent reason why it should not be re-adopted) and the analyses are now depicted by a two dimensional display. Analysis of the present group of abstracts revealed four cases in which the two dimensional representation was inadequate and a need was felt to move into

three dimensions. These four examples are given below as they represent a feature of detailed analysis that has not previously been encountered.

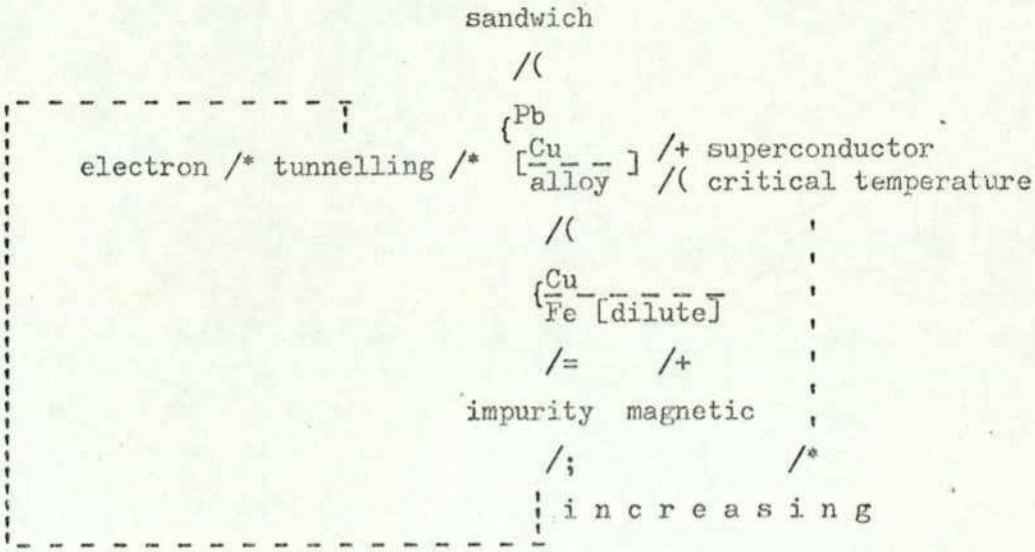
34/0357:



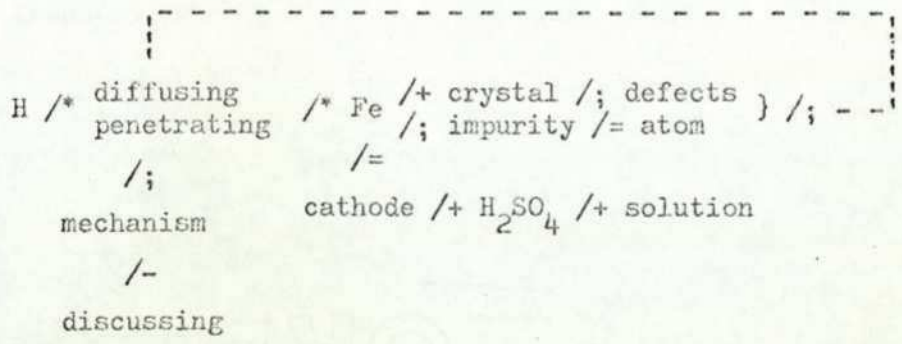
34/0409:



33/1792:



34/0565:



3.1.4 Indexing and preparation times

One of the criticisms levied at relational indexing by people who have not become fully acquainted with the system, is that it takes a long time to index items with this technique and that, by implication, the input costs would be higher than those for other controlled language systems. In order to investigate this aspect, the various functions in preparing the relationally indexed card index were timed over a period. The various average times are set out in Table 3.

TABLE 3

Average times to perform operations in preparing the card index

Time to index	5.28 mins/abstract
Typing time (for master)	3.81 mins/abstract
Duplicating time	0.98 mins/abstract
Card heading time	4.16 mins/abstract
	0.42 mins/card
Filing time	2.39 mins/abstract
	0.24 mins/card
Total clerical time	11.34 mins/abstract
Total time (indexing + clerical)	16.62 mins/abstract

The sample sizes taken for calculating these average times were 957 abstracts for the indexing operation and 832 abstracts for the various clerical operations. The average

time of 5.28 minutes for indexing an abstract compares favourably with other systems [e.g. 57, 60]. The clerical operations do not take an excessively long time with an average of 11.34 minutes per abstract. All these times were calculated when the operators were reasonably proficient and the most difficult part of the clerical operations is typing the spirit duplicating master. Seeing the best layout and appropriate spacing in order that the indexed version will fit into a 6 inch by 4 inch card is a skill which takes a little time to acquire, but the degree of difficulty depends on the way in which the indexer lays out the work for the typist. The actual times taken for the clerical operations will depend on the depth of indexing used and, as the depth is increased, the volume of work at each stage is increased similarly. In this project an average of 9.81 cards per abstract were produced in order to give a suitable number of entry points to the card index. Obviously if a large system, based on relational indexing, were being set up, this method of preparing a card index would involve a considerable amount of clerical labour. (As has been previously stated in Section 2.2, a total of some 32,000 cards were created for this research.) However, in a computer-based system the clerical involvement could be considerably reduced.

3.1.5 Exhaustivity of indexing

The quantitative assessment of exhaustivity and depth of indexing has not been resolved as yet. The exhaustivity of indexing of an abstract depends on the size of the abstract. Obviously a title can be exhaustively indexed and still have fewer

index terms than an abstract which is indexed in lesser detail. However, the number of index terms or concepts per document is about the only measure that can be adopted at present, although this will be influenced by the type and control of the vocabulary used in different projects.

In this project the average number of indexed concepts per document was 12.1. The "exhaustivity" for each section was also determined as follows:

Section	No. of concepts per document
1.2	12.5
1.4	14.7
3.2	10.8
3.3	11.4
3.4	12.5

The reasons for this variation with sections could be due to several factors. Variation in the size of the abstract is considered unlikely, as abstracts of different sizes were observed in all sections. It may be that section 1.4 warrants more detailed indexing than say section 3.2, or that the abstracts on "structural hardening" contain more significant terms than those on "physical properties". Or it may be a simple explanation that work on structural hardening is usually conducted on complex alloys, whereas the physical properties are usually measured of elements or simple compounds.

Thus the significance of an average of 12.1 terms per document is not particularly clear. In discussing such aspects, Sparck Jones [93] quotes figures for the Cranfield, INSPEC and College of Librarianship Wales projects as 32, 12.2 and 7.9 terms per document respectively. However, a more detailed comparison of the four document collections would be necessary before any conclusion could be drawn.

3.1.6 Breadth of indexing

The breadth of indexing terms, or the number of postings made to the terms, was not studied in this project, although its effects were implicitly acknowledged. Te Nuyl [95] has shown that terms with relatively few postings are more efficient in retrieval. Sparck Jones [93] has proposed that specificity should be defined statistically so that specific terms are those with low posting frequency. Svenonius [94] has also considered breadth of indexing and both she and Sparck Jones conclude that a mixture of specific and non-specific (narrow and broad) is best for retrieval, depending on whether high recall or high precision is required.

As far as relational indexing is concerned, the breadth should not be so critical as the structural pattern plus the browsing techniques give greater flexibility for retrieval. In fact broad terms, such as "alloy" and "crystal", were not even used as index entry points, although they were of course included in the structure for matching.

3.1.7 Generic browsing

The matching strategies for the project were discussed in Section 2.6. The third strategy mentioned was generic browsing and this technique is designed to overcome one aspect of the difference in depth and specificity of indexing between the profiles and the abstracts. In order to apply the technique one or more terms in the profile are replaced by a more specific or a more generic term. To evaluate the browsing method adequately it is necessary to have a fully developed classification with concepts completely specified within their hierarchies. It would be desirable that the levels were quantified as one step in one hierarchy may be equivalent to two or three steps in another hierarchy. However, no classification which embodies this latter aspect is known. Also the concepts were not arranged in a full classification for this project, although they were controlled to the extent of being placed in a standard list and the generic browsing was conducted by implied classification.

Generic browsing can occur with one term and it may go up or down the hierarchy. The principle may also be extended to two or more terms in the profile, but if this is so, the browsing profile can become widely divergent from the original profile within a very few browsing steps.

The indexing of the abstracts for the card index was fairly specific and very little generic posting was used, unless the generic term was embodied within the abstract. In contrast

the profiles were very often couched in general terms and this meant that the generic browsing was one of the most frequently adopted browsing techniques and also that it was mainly conducted in a downwards direction.

Because of the high frequency of occurrence, the technique was given detailed study for the November and December issues only, and the results of this analysis are given in Table 4.

TABLE 4

Analysis of the generic browsing strategy

	No. of instances giving selected abstracts	No. of instances giving non-selected abstracts	Total
November			
Upward browsing	15 (46%)	18	33
Downward browsing	79 (67%)	39	118
Total	94 (62%)	57	151
December			
Upward browsing	13 (65%)	7	20
Downward browsing	70 (82%)	15	85
Total	83 (79%)	22	105

From this it can be seen that in general terms, browsing down a classification to more specific concepts is more likely to give relevant papers than browsing up to more generic terms. In this project, this is considered to be a reflection of the way in which the profiles for the SDI situation were expressed. This may also

be true of retrospective searching situations, as was found in the first City University project [32], but it is probably a function of the balance of specificity of indexing between profiles (questions) and abstracts. (Examples of generic and other browsing techniques used are given in Appendix IV.)

3.1.8 Conceptual browsing

This technique was also used in the first City University project. It is more sophisticated than generic browsing, but is not so well developed. The essential basis of the method is that various combinations of concepts and relations can be used to express ideas which are sufficiently similar in nature and meaning to be regarded as equivalent for searching purposes. A simple example of this type of browsing would be that "studying crystal structure with an X-ray diffractometer", which would be indexed as

crystal /(structure /- studying /; X-ray diffractometer

might be considered as a relevant answer to a profile which was concerned with "uses of an X-ray diffractometer", which would be indexed as

X-ray diffractometer /; uses.

Thus a "searching pattern" for the latter question might be

A /- B /; X-ray diffractometer

where A and B represent any concepts. Thus it involves not

merely a change in concept(s) in a hierarchical structure, but possibly a complete change in one or more concepts allied with a possible change in relations and in structure.

Apart from a few specific examples, such as the one given above, this technique has not been thoroughly investigated and there are no standard rules which can be adopted. Thus at present it tends to be a rather intuitive method of selecting possibly relevant abstracts.

Its application in this project was not as extensive as that for generic browsing, nor were the results obtained with its use as successful overall. In fact it was applied in a total of 301 matching operations throughout the six months and of these, 178 (59.1%) were successful in selecting relevant items. (It is intended to make the analysis of these 301 matching operations the subject of a later project.)

3.1.9 Condensations

The technique known as condensations or logical jumps has been previously mentioned in Section 2.6 and it has been fully described elsewhere [36, 38, 82, 101]. It is a technique which has been developed to overcome differences in exhaustivity, rather than specificity, of indexing and the essence of the approach is that in certain situations the middle concept and one relation may be omitted out of a string

of three concepts and two relations, i.e. A / B / C may condense to A / C, where A, B and C are concepts and the stroke (/) represents any relation. A series of 32 empirical rules have been developed which specify which relation out of a combination of two relations is omitted.

There are certain variations on this simple basis. Simultaneous multiple condensations can occur which result in more than one concept, with an equal number of relations, being omitted, e.g. A / B / C / D / E may condense to A / E. These multiple jumps are equivalent to chains of justifiable single jumps. Jump-backs are another variation in which the condensation occurs in an angular structure, for example

```
      A                    A
      /                    /
C / B      may condense to   C      or  A / C,
```

```
B / C                C                C / A.
   /                 /
  A                  A
```

or B / C may condense to C or C / A.

The condensations applied in this research were analysed. The three term string condensations of the simple type were the most frequently occurring, and the analysis is shown in Table 5.

A number of double, triple and quadruple string type condensations were also used. A double jump is the omission of two concepts with two relations, a triple jump the omission of three concepts with three relations and a quadruple jump the omission of four concepts with four relations. The analysis of

TABLE 5

Analysis of simple condensations

Rule	First relation	Second relation	Omitted relation	No. of occurrences	Relevant abstracts	Non relevant abstracts
2	any	same	either	53	39	14
4	/*	/-	/*	6	2	4
5	/+	/;	/+	25	22	3
8	/;	/(	/(	2	2	-
10	/(	/+	/(	6	3	3
11	/(	/-	/(	7	1	6
12	/(	/;	/(	10	7	3
13	/(	/*	/(	15	12	3
14	/(	/:	/(	4	3	1
15	any (or /=)	/= (or any)	/=	17	11	6
20	/*	/;	/*	8	5	3
21	/+	/*	/+	3	3	-
22	/+	/-	/+	11	5	6
23	/+	/(	/(	24	20	4
28	/:	/*	/:	22	12	10
31	/:	/;	/:	1	1	-
			TOTAL	214	148 (69%)	66 (31%)

these condensations is given in Table 6.

TABLE 6

Analysis of double, triple and quadruple condensations

	No. of occurrences	Relevant abstracts	Non relevant abstracts
Double jumps	31	18	13
Triple jumps	6	5	1
Quadruple jumps	3	2	1
TOTAL	40	25 (62.5%)	15 (37.5%)

Rules of the same type as those devised empirically for single jumps could be devised for multiple jumps, although the number of permutations of the nine relations would lead to an impossibly large number. Thus for quadruple jumps the number of rules would be $\frac{9!}{5!} = 3024$. A simpler way of handling such problems is to assign a "strength" to each relation so that a "weaker" relation will give way to a "stronger" one in a condensation. The relations can be listed in an order of strength from the 32 empirical rules for single jumps and this has been suggested [38]. However, by studying the multiple jumps, a similar order can also be devised, although a large number of examples would be needed to accurately define a complete listing including all nine relations. From the double jumps the following order was prepared:

/- /; /* /+ /:

/(

/=

strength ⟵⟶ weakness

The triple and quadruple jumps did not contradict this order, but the number of occurrences was not large enough to give such a complete diagram.

A total of 23 jumpbacks were also observed, 12 of which produced relevant abstracts and 11 of which produced non-relevant abstracts. Fifteen of the total were single jumpbacks involving the omission of one concept with one relation, for which nine empirical rules have been suggested [38]. The structure of these jumpbacks was either

A $/_1$ B
$/_2$
C

or

B
$/_1$
C $/_2$ A

and the occurrences are shown in Table 7.

Of the eight multiple jumpbacks, two produced relevant abstracts and six irrelevant abstracts. A "strength" spectrum can be deduced from the eight occurrences, which is as follows:

/; /* /+

/(

/-

/=

strength ⟵⟶ weakness

The spectrum is similar to that obtained for the multiple string condensations, except that it is not so fully developed and the position of the reaction relation is changed.

TABLE 7

Occurrences of single jumpbacks

Rule	Relation 1 ($/_1$)	Relation 2 ($/_2$)	Omitted relation	No. of occurrences	Relevant abstracts	Non-relevant abstracts
2	/+	/(	/(	3	-	3
3	/+	/;	/+	1	-	1
5	/;	/(	/(	2	1	1
new	/(	/+	/+	2	2	-
new	/(	/=	/=	6	6	-
new	/+	/=	/=	1	1	-
		TOTAL		15	10 (67%)	5 (33%)

The analyses given above include the majority of the condensations occurring during the six months of the project. However, some had to be omitted because they were apparently

caused by erroneous indexing and hence were not true condensations. These results tend to give a false impression of the value of this technique or the correctness of the rules because of the relatively high number of condensations which resulted in abstracts which the user did not consider relevant. In fact the number of incorrect condensations was very small (see Section 3.2.16) and the reasons for non-selection by the user are many and varied. Some idea of the number of non-relevant abstracts that are likely to be selected by any one searching strategy can be obtained by comparing the results with those of the "match plus" strategy. This would give a quantitative estimate of the proportion of the data base which is likely to be considered irrelevant by the user on the basis of additional concepts and relations over and above those specifically stated in his profile. In the vast majority of cases it is the abstract which is condensed to meet the requirements of the profile and thus it is perhaps fair to conclude that the concept omitted by the condensation may be sufficiently significant as to render the abstract not relevant.

3.1.10 Related browsing

Generic browsing consists of replacing one or more concepts in the profile by concepts either higher up or lower down a hierarchy. Because the new concept is of the same type as the one being replaced, the relations between the new concept and any others remain the same.

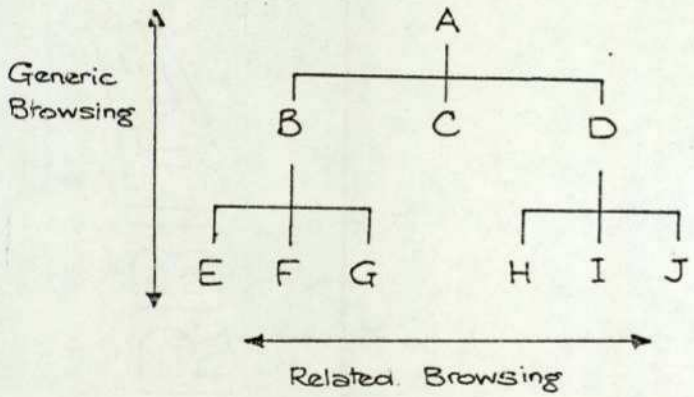
Related browsing is very similar in operation; in this technique one or more concepts in a profile may be replaced by a related concept, which is not in a direct genus-species relationship with the original concept. Thus generic browsing can be thought of as a vertical movement, whereas related browsing would be a horizontal movement. The differences can perhaps be more clearly illustrated in relation to an outline classification structure such as Farradane's [31], as depicted in Fig. 6. Generic browsing is restricted to a class structure, whereas related browsing may be contained within a class structure (browsing in array) or it may traverse concept type boundaries, for example, an abstract concept may be replaced by an activity concept. This technique has not been used before in relationally indexed retrieval systems and hence there are no predctermined rules which can be directly used.

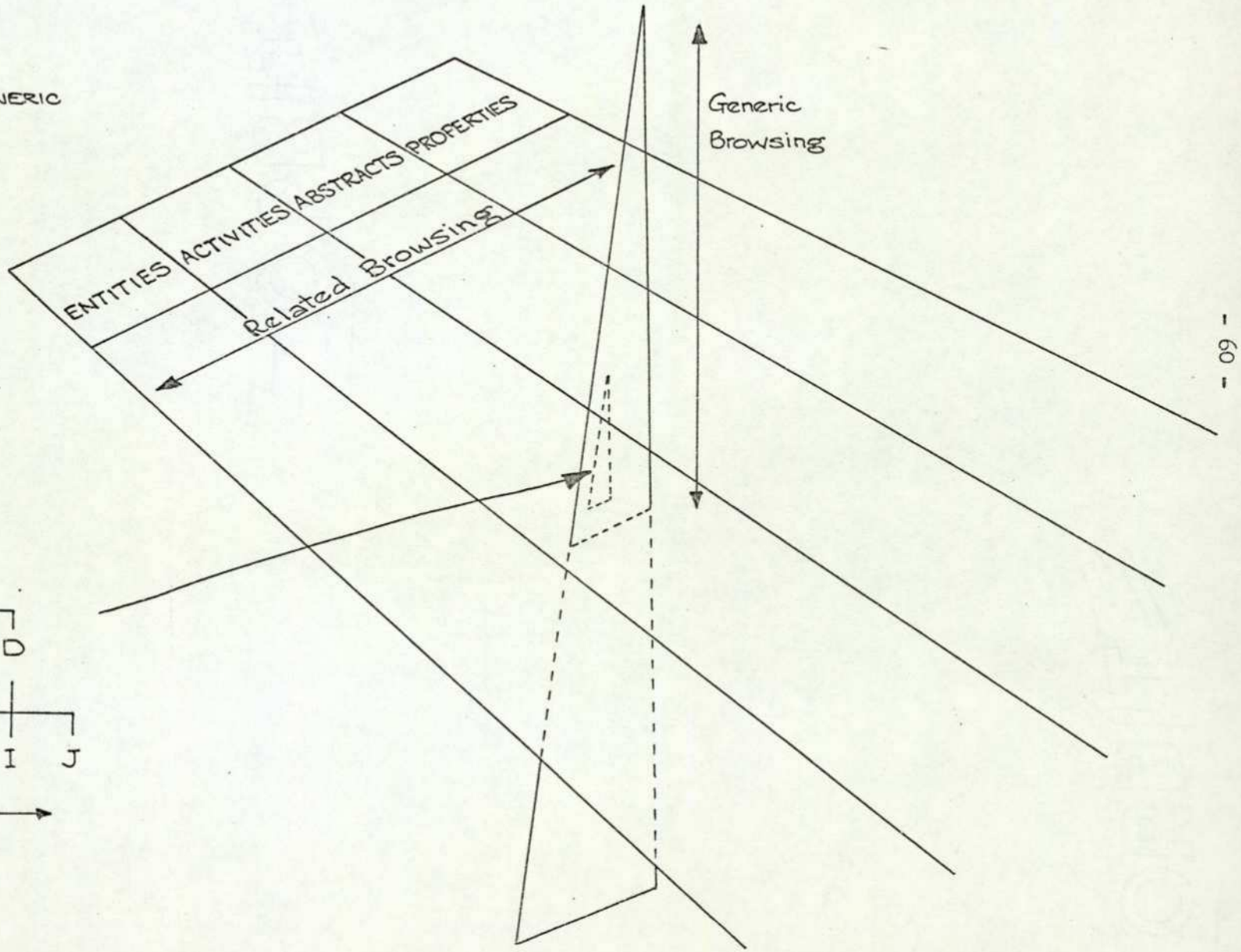
Related browsing is a technique which must obviously be used with care, as there is greater scope for diverging from the original profile with perhaps one or two substitute concepts. As with the other searching strategies it is useful in overcoming some of the difficulties imposed by differences in depth and exhaustivity of indexing between questions and the data base. However, the extent to which it is likely to be successful in retrieving abstracts which would not otherwise be selected depends on the precision with which the user states his question. Another aspect of this technique which may lead to erroneous selections is that not only the concept, but also the relation,

FIGURE 6

THE DIFFERENCE BETWEEN GENERIC

AND RELATED BROWSING





may have to be changed. As the browsing is conducted not up and down hierarchies, but across them, it is inevitable that changes should occur. For example, a user who is interested in "the crystal form of an alloy", indexed as

alloy /+ crystal,

may also be interested in a paper on the "crystallization of the alloy", indexed as

alloy /* crystallizing.

In this case the original concept, "crystal", is an abstract (form) concept, linked to "alloy" by a dimensional relation, whereas the activity "crystallizing" is linked by the self-activity relation.

Thus in a sense related browsing is a sort of halfway house between generic and conceptual browsing. However, developing rules for such a searching strategy will need careful consideration. If the basic controlled vocabulary list is in the form of Farradane's outline classification, then browsing could take place from say "properties" across to equivalent levels in "abstracts", "activities" and "entities" and also to heterogeneous concepts. But for rules to be established, the concepts would have to be fully structured into this classification format. Similarly in a pure hierarchical classification, related concepts would be those within the same array and possibly also in other parallel arrays. Whether or not the "related terms" in a

conventional thesaurus would provide the type of browsing that is required would depend in part on the way in which the thesaurus was constructed and compiler's definition of "related term". This is an area of classification which has received little thought and it is worth noting that the Classification Research Group are now beginning to turn their attention to rules for obtaining related terms.

Because of present uncertainties, the technique of related browsing was sparingly used in the present research and on those occasions when it was used, the methodology had to be mainly intuitive. In fact the greater proportion (56%) of its applications led to the selection of non-relevant items, but it is considered that the number of cases were insufficiently large to enable any valid conclusions to be drawn. For example, it may be that the loss in precision with this technique would outweigh any gains in recall, but a more complete investigation would be necessary before such a conclusion could be reached.

The related browsing occurrences were analysed according to the types of concepts involved and the resultant analysis is shown in Table 8. The most frequent instances were between concepts of similar types, e.g. property to property, or browsing in array. The main point to emerge from this is that perhaps the browsing technique cannot be considered in isolation from the remainder of the profile and abstract, as these may in fact contain more significant concepts than those

TABLE 8

Analysis of related browsing occurrences

Concept from profile	Concept from abstract	Relevant abstracts	Non-relevant abstracts
Entity	Entity	1	1
Entity	Activity	1	1
Activity	Activity	5	12
Activity	Abstract	1	-
Activity	Property	5	2
Abstract	Activity	1	1
Abstract	Abstract	3	6
Abstract	Property	1	-
Abstract	Heterogeneous	-	1
Property	Entity	1	-
Property	Activity	6	10
Property	Abstract	1	2
Property	Property	7	9
Property	Heterogeneous	1	-
Heterogeneous	Abstract	1	-
		35	45

involved in the browsing operation. For example, "hardness" in a profile was replaced by "hardening" on six occasions; on two of these occasions the abstract was considered relevant by the user, but on four occasions the user rejected the abstract. Similar situations were noted with the other browsing techniques.

3.1.11 Analysis of selection strategies

The usefulness of different strategies and their effectiveness in improving recall and precision can be studied in various ways. The simplest way of studying their effects would be to make a selection on the basis of complete matches between profiles and abstracts and calculate recall, precision and fallout. The selections could then be repeated using the match plus strategy and the ratios re-calculated. Such a procedure could then be applied to each strategy and their individual ratios, together with their combined effects, calculated and compared.

There were two main factors which militated against such a methodology. One of these was time. Performing the six strategical matching operations for 43 profiles for each issue of the abstracts journal would have taken up far more time than was available to the researcher. The other factor against such an approach for an SDI system is that only a few, if any, items would be selected by any one strategy for each profile per issue. The problem of small values of R, C and L, compared with N, results

in very small changes in these parameters giving disproportionately large percentage changes. Hence in this research the various strategies were all applied during the one monthly selection operation. The strategy used was recorded in each case and hence their respective effects have been ascertained subsequently.

An analysis of the use of the different searching strategies is given in Table 9. It is perhaps not surprising that the majority of both relevant and non-relevant abstracts were selected by the "match plus" strategy and that "generic browsing" should produce the next largest volumes in both categories.

For the majority of the profiles the match plus strategy was most frequently employed in the selection process and the analysis of which strategy was the most used for the various profiles is given in Table 10. It is a reflection of the way in which some of the profiles were constructed that it was necessary to adopt browsing techniques in the majority of selections, but is also perhaps an indication of the necessity for adopting such strategies for effective retrieval.

The actual figures and comparisons for the effectiveness of the searching strategies in terms of recall and precision are given and discussed in Section 3.2.17.

TABLE 9

Monthly analysis of searching strategies

Searching strategy	Number of selections per issue							
	7007	7008	7009	7010	7011	7012	Total	%
Relevant abstracts								
Match plus	138	98	90	115	207	186	834	51.6
Generic browsing	99	32	69	38	94	83	415	25.7
Conceptual browsing	56	21	19	18	31	33	178	11.0
Condensations	35	20	25	15	33	24	152	9.4
Related browsing	9	7	4	3	10	5	38	2.3
Sub-total	337	178	207	189	375	331	1617	100.0
Non-relevant abstracts								
Match plus	165	74	31	26	62	65	423	43.0
Generic browsing	89	29	20	21	57	22	238	24.2
Conceptual browsing	36	15	13	16	29	14	123	12.5
Condensations	46	32	17	15	33	9	152	15.4
Related browsing	27	4	5	2	3	7	48	4.9
Sub-total	363	154	86	80	184	117	984	100.0
TOTAL	700	332	293	269	559	448	2601	-

TABLE 10

Analysis of selection strategy most used by profiles

	Number of profiles using strategy most frequently	
Strategy	Relevant abstracts	Non-relevant abstracts
Match plus	24	21
Generic browsing	10	11
Conceptual browsing	5	2
Condensations	2	4
Related browsing	-	-
	41	38

N B The totals do not add up to the 43 profiles used because some profiles used two or more strategies equally.

3.2 Evaluation

3.2.1 Primary parameters

The use of the 2 x 2 contingency table as the basis of evaluation involved the determination of the four primary parameters, R, C, L and N. The initial determination of these four parameters was relatively straight-forward, using the methodology described in Section 2.7 to obtain a value for C, which usually presents the largest problems. As previously described, the documents in the sets (C - R) and (L - R) were re-examined and, if necessary, these documents were re-submitted to the user. Thus the values of both C and R may be altered, and this provides the second set of values, C' and R'. The errors that caused the differences between C, C' and R, R' were called "User selection errors" and these were regarded as "acceptable" errors in so far as they were directly attributable to the human foibles of the users.

There were a number of other "acceptable" errors, also attributable to human factors, some due to the users, others due to the selector or other parts of the "system". These errors were:

User	System
User profile	Selecting
Sixth sense	Typing
Language	Technical knowledge
User's idiosyncrasies	

They are discussed in greater detail in Section 3.2.16, although their interpretations are reasonably self-explanatory. Allowance was made for these errors, which can affect R, C and L, so that these modified parameters provide a third set of values, R", C" and L".

The value of N is another debateable point, which is discussed in Section 3.2.10.

3.2.2 Calculated parameters

The results were calculated from these four primary parameters to yield recall, precision, fallout, generality, Q and (M x P) values. The results were grouped into sets, each of which was designated "XYZ". For sets 071 to 123,

XY represents the month of issue, e.g. 07 = July, 12 = December

Z represents the values of R, C and L used, i.e.

1 = R, C, L; 2 = R', C', L; 3 = R", C", L".

Thus there were three sets of data for each of the six monthly issues. These results were then cumulated into sets 708 to 712, where 708 represents the sum of July and August, 709 the sum of July, August and September, etc. and 712 the total of the six months. These cumulations were calculated only for the final values of the primary parameters, i.e. R", C" and L", as these results were considered to be the most pertinent.

Within each set the relevance assessments were included. The users assigned each selected item to relevance 1 ("of definite interest") or relevance 2 ("of possible interest"). Separate results were calculated for relevance 1, relevance 2 and relevance (1 + 2).

These results were plotted as scatter diagrams on recall - precision and recall - fallout axes and are shown in Figs. 7 to 14. (All graphical representations of results are placed together at the end of Section 3.2, following page 125.) The arithmetic means of the various fractions (or percentages) are tabulated in Table 11. These values all refer to combined relevance (1 + 2). (The individual results for relevance 1 and relevance 2 are shown in Figs. 50 and 51 and are discussed in Section 3.2.13.) The recall, precision and fallout ratios for various sets are given in Appendix II.

3.2.3 Effect of parameter values

Recall - precision and recall - fallout diagrams for sets 071, 072 and 073 are shown in Figs. 7, 8 and 9. For all three sets the individual results show a fairly high degree of scatter over the whole diagram, a feature which is becoming anticipated in ISR system testing. It is to be expected that the results would improve in passing from set 071, through 072 to 073, as they include increasing allowance for errors. However, in the recall - precision diagrams there is a tendency, which is

TABLE 11

Arithmetic means of results (for relevance (1 + 2))

Set	Recall	Precision	Fallout
071	0.3827	0.3019	0.01631
072	0.4711	0.4422	0.01203
073	0.6896	0.7220	0.00410
081	0.3639	0.3222	0.01290
082	0.5221	0.5244	0.00920
083	0.7137	0.7109	0.00479
091	0.4497	0.3993	0.01321
092	0.6428	0.6337	0.00742
093	0.7774	0.8729	0.00302
101	0.4798	0.4401	0.01177
102	0.6616	0.6855	0.00687
103	0.7865	0.8475	0.00281
111	0.4570	0.3346	0.01453
112	0.8001	0.6244	0.00754
113	0.8009	0.7634	0.00439
121	0.4843	0.4601	0.01360
122	0.6592	0.7300	0.00619
123	0.8159	0.8689	0.00274
708	0.7099	0.6966	0.00419
709	0.7192	0.7199	0.00392
710	0.7266	0.7372	0.00373
711	0.7409	0.7450	0.00390
712	0.7591	0.7737	0.00368

most striking in Fig. 8, for the results to group round the positive diagonal. This tendency would seem to indicate that, depending on the actual question, the system produces both high recall and precision or vice versa low recall and precision. In other words the recall and precision tend to be of similar values, rather than one parameter being high at the expense of the other.

Comparison of averages for sets XY1, XY2 and XY3 for each month (Table 11) shows similar trends in each case.

3.2.4 Variation over the six months

Figs. 9 to 14 inclusive depict the recall - precision and recall - fallout diagrams for sets 073, 083, 093, 103, 113 and 123, i.e. the final figures calculated from R", C" and L", for each of the six issues of Metals Abstracts.

For both types of diagrams the results show an apparent improvement with time, although it is difficult to compare scatter diagrams as such. In the recall - precision diagrams, the tendency to group round the positive diagonal is still present, although perhaps the trend for the results to group in the top right hand corner is more marked. Although the average of results with such degrees of scatter are of relatively little value, nevertheless the average values shown in Table 11 enable the trends to be more easily seen. These show an overall improvement with time, although the trend is not smooth for

precision and fallout.

In a six month run of an experiment such as this, the results for each issue should be similar if all other parameters remain unchanged. Therefore the overall improvement in results from July to December, must be due to some disturbance in the system. This may be a direct result of profile modifications of which there were a reasonable number throughout the six months, owing to changing user interests and inaccurate expression or interpretation of the original interests. A further possible interpretation could be that the users were becoming less critical of the system selection as the project proceeded.

The monthly averages are depicted against time in Figs. 15, 16 and 17, from which it can be seen that although recall shows a reasonably steady improvement, both precision and fallout are more erratic. Again the movement of the average values on recall - precision and recall - fallout diagrams is zigzag rather than a straight line (Fig. 18).

3.2.5 Cumulation of results

The primary parameters for each user were cumulated and ratios of the type $\frac{R_{July} + R_{August}}{C_{July} + C_{August}}$ to $\frac{R_{July} + \ldots + R_{December}}{C_{July} + \ldots + C_{December}}$ were calculated for recall, precision and fallout. The final cumulation for the total six months (set 712) gave scatter diagrams

of the type shown in Fig. 19 and the average values for the five cumulative sets are included in Table 11. It is not surprising that these averages, with the exception of fallout, should steadily increase through the duration of the project, and in fact they reflect a smoothing effect over the results for the individual months. The cumulative averages are plotted in Figs. 15 to 18 inclusive.

The effect of averaging is shown in the scatter diagram (Fig. 19) by a more condensed series of points, fewer of which reach the edges of the diagram, i.e. recall or precision equals 1.0 and fallout equals 0.

3.2.6 Overall performance over six months

The performance during the six months duration of the project has been discussed in Section 3.2.5, the variation being studied in terms of recall - precision and recall - fallout diagrams and the average values of recall, precision and fallout. The comparison of diagrams is not wholly satisfactory, nor is the comparison of averages of ratios. An overall performance measure would permit a more uniform comparative analysis to be made.

Various overall performance measures were discussed in Section 2.8 and of these the Q value was not immediately eliminated and the product (M x P) retained, although both have their disadvantages. Referring back to the 2 x 2 contingency

table given in Section 2.8 (page 31), the Q value is defined as

$$Q = \frac{R(N - C - L + R) - (C - R)(L - R)}{R(N - C - L + R) + (C - R)(L - R)}$$

which simplifies to

$$Q = \frac{M - F}{M + F - 2MF}, \text{ where } M = \text{recall and } F = \text{fallout.}$$

This statistic expresses the degree of association between the "selection power" and "acceptability" of papers achieved by the system. The main disadvantages of it as an overall measure are that it is not sensitive to small changes in parameters and it is restricted in handling terminal conditions. Thus Q = 1 when either F = 0 or M = 1 and Q = -1 when either F = 1 or M = 0, as shown in the recall - fallout diagram of Fig. 20.

There are various circumstances which may give rise to the situations mentioned above. For example, if C = x, L = y, R = 0, then M = 0 and hence Q = -1, irrespective of whether x is 1 or 200. Similarly, if C = x, L = y, R = y, then F = 0 and hence Q = 1, irrespective of whether $x \doteq y$ or $x \gg y$. Although a system which gave points in the lower right hand half of this diagram could not be considered as operating efficiently, there is a small area near the origin (enclosed by a dotted line) in which seemingly bad results may be due to only relatively small errors in system performance.

Although the ultimate objective of a system is the top left hand corner, viz. M = 1, F = 0, the Q value makes no

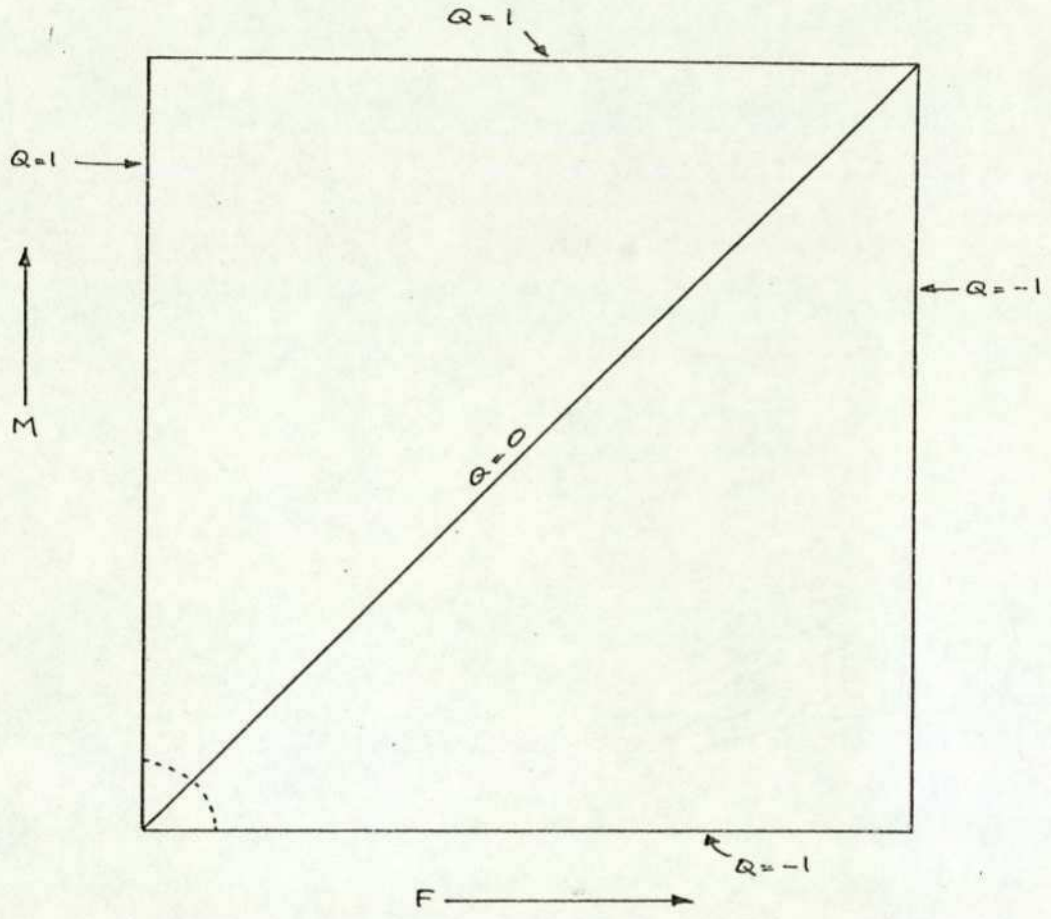


FIGURE 20

General recall - fallout

diagram, showing Q values

distinction between this point and any other which falls on the upper or left hand boundaries of the diagram.

The Q values embodies the same disadvantages as do other measures which incorporate R, L, C and N, namely that, as R, L and C are generally considerably smaller in value than N, the value of N dominates the measure. The product (M x P) is not similarly dominated, and consequently is subject to the other criticism of ISR system testing measures, namely that it does not incorporate all parameters of the 2 x 2 table. On a recall - precision diagram the (M x P) lines are hyperbolae across the diagram as shown in Fig. 21. This measure embodies an implicit acceptance of an inverse relationship between recall and precision, which has been recently qualified by Cleverdon [17]. Nevertheless, if it is assumed that one of the ultimate objectives of an ISR system is the attainment of both recall and precision equal to unity, then this value gives some measure of the attainment of this objective. One advantage of the product is that, for the same results, it has a wider range of values than the Q value. Therefore both values are used for the overall assessment of profiles.

The Q values for each profile during the six month period are given in Table 12, from which it can be seen that their performance was not very clearly defined. In only six cases (profiles 1, 30, 37, 54, 56 and 66) did the profiles show a reasonably consistent improvement during the period. In 18 cases the performance was virtually static and in a further 18 cases it was oscillatory. In only one case (profile 10) did the performance

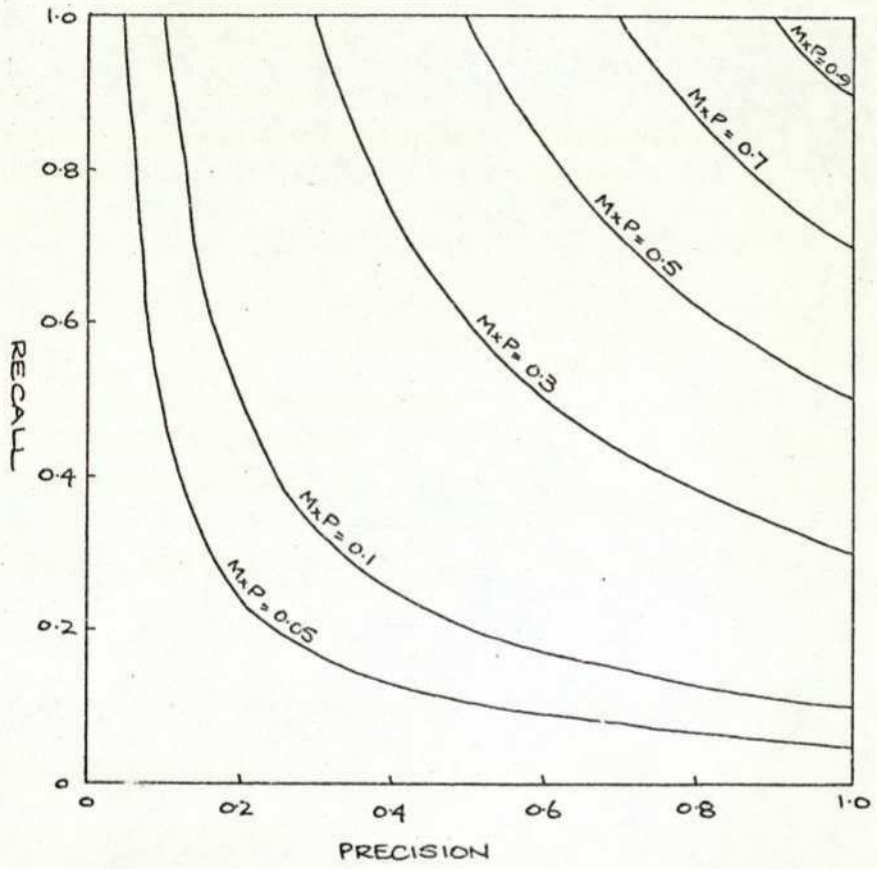


FIGURE 21

Recall - precision diagram

showing lines of constant (M x P)

TABLE 12

Monthly Q values

Profile Number	7007	7008	7009	7010	7011	7012
1	-1.000	0/0	0/0	0/0	1.000	1.000
3	1.000	1.000	0/0	1.000	1.000	1.000
6	0/0	0/0	0/0	0/0	0/0	0/0
7	0.996	0.993	1.000	1.000	0.998	0.998
9	0.996	0/0	1.000	1.000	0.998	0.999
10	1.000	0.998	0.979	1.000	0.998	0.997
12	1.000	1.000	1.000	1.000	1.000	1.000
13	0.981	0/0	1.000	1.000	0.997	1.000
14	1.000	1.000	0.999	0.997	1.000	1.000
21	0.999	1.000	1.000	0.998	0.999	1.000
22	1.000	0.982	1.000	1.000	1.000	0.983
24	1.000	1.000	1.000	1.000	1.000	1.000
26	0.998	0.994	0.996	0.981	0.995	0.998
27	0.997	0.986	1.000	0.994	0.971	0.996
29	1.000	0.995	0.994	0.991	0.996	0.998
30	0.986	0.989	1.000	0.993	0.999	1.000
31	0.999	0.996	1.000	1.000	1.000	0.996
32	1.000	1.000	1.000	1.000	1.000	1.000
34	0.980	0.960	1.000	0.991	0.998	0.999
35	1.000	0/0	0/0	0/0	1.000	0/0
36	1.000	1.000	1.000	1.000	1.000	1.000
37	1.000	1.000	1.000	1.000	0.996	1.000
40	0.967	0.992	0/0	1.000	0.999	1.000
41	0.986	1.000	1.000	0.986	1.000	1.000
42	0.999	0.991	1.000	1.000	0.999	0.999
43	1.000	0.998	1.000	1.000	1.000	1.000
47	1.000	1.000	1.000	1.000	0.999	1.000
51	0/0	1.000	0.989	0.998	1.000	1.000
53	0.989	0.996	0.993	0.997	0.997	0.995
54	0.980	0.997	0.995	1.000	0.995	0.999
56	0.997	1.000	1.000	1.000	1.000	1.000
57	0.986	0.988	1.000	1.000	0.980	0.993
58	0.997	1.000	1.000	0.995	1.000	1.000
59	0.999	1.000	1.000	1.000	1.000	1.000
60	0.987	0.996	0.999	0.986	0.967	1.000
61	1.000	1.000	1.000	1.000	1.000	0.999
62	1.000	0.977	1.000	0.999	1.000	0.998
63	0.998	1.000	1.000	1.000	1.000	1.000
64	0.967	0.976	1.000	0.998	0.989	1.000
65	0.912	1.000	1.000	0.997	0.960	0.966
66	1.000	0.973	0.995	0.997	0.998	1.000
67	1.000	0/0	0/0	1.000	1.000	0/0
69	1.000	0.991	0.997	0.993	0.998	0.997

deteriorate during the period.

For set 712 (that is the combined results over the six months) the Q values of the profiles are given in Table 13, together with the (M x P) values. This demonstrates the differences in the ranges, which for Q was from 0.9804 to 1.0000 and for (M x P) was from 0.212 to 0.930. The comparative usefulness of the values for the ordering of the results is debatable, but it is not surprising that they should be different, although there is some correlation between the two (Spearman's coefficient of correlation, r = 0.73 and the partial tetrachloric coefficient = 0.81).

3.2.7 The recall - fallout graph

The recall - fallout diagrams are given in Figs. 7 to 14 and the Q value is discussed in the previous section. The results, as shown on recall - fallout diagrams, have not been discussed.

Curves of constant Q can be drawn on a recall - fallout diagram and they have a hyperbolic nature (see Fig. 22). Whether or not these data fit these curves remains a mute question. In the first City University project [32] the results were such that the Q value was very frequently equal to unity and hence the few other values were inadequate to draw conclusions. Again, in the present research the results did not appear to group round discrete Q values, as was seen from Tables 12 and 13 and Figs. 7 to 14.

TABLE 13

Correlation of Q and (M x P) (set 712)

Profile Number	Q	Rank order (Q)	(M x P)	Rank order (M x P)	Difference (D)	D^2
1	0.9995	11	0.457	32	21	441
3	0.9999	5	0.790	5	0	0
7	0.9963	32	0.499	29	3	9
9	0.9976	24	0.496	30	6	36
10	0.9975	25	0.572	23	2	4
12	0.9999	3	0.930	1	2	4
13	0.9909	38	0.315	39	1	1
14	0.9992	16	0.612	20	4	16
21	0.9994	13	0.748	10	3	9
22	0.9965	31	0.445	34	3	9
24	0.9999	5	0.910	2	3	9
26	0.9941	33	0.556	24	9	81
27	0.9915	37	0.447	33	4	16
29	0.9968	29	0.728	13	16	256
30	0.9969	28	0.412	36	8	64
31	0.9991	17	0.711	16	1	1
32	0.9997	9	0.721	14	5	25
34	0.9922	36	0.531	26	10	100
35	0.9999	5	0.766	8	3	9
36	1.0000	1	0.756	9	8	64
37	0.9996	10	0.794	4	6	36
40	0.9967	30	0.503	28	2	4
41	0.9970	27	0.402	37	10	100
42	0.9989	21	0.678	18	3	9
43	0.9998	8	0.782	6	2	4
47	0.9990	20	0.732	12	8	64
51	0.9993	15	0.593	22	7	49
53	0.9935	34	0.552	25	9	81
54	0.9934	35	0.430	35	0	0
56	0.9995	11	0.742	11	0	0
57	0.9847	41	0.290	40	1	1
58	0.9991	17	0.519	27	10	100
59	0.9991	17	0.696	17	0	0
60	0.9898	39	0.468	31	8	64
61	0.9999	4	0.868	3	1	1
62	0.9988	22	0.776	7	15	225
63	0.9994	13	0.713	15	2	4
64	0.9876	40	0.266	41	1	1
65	0.9804	42	0.212	42	0	0
66	0.9980	23	0.625	19	4	16
67	1.0000	1	0.353	38	37	1369
69	0.9973	26	0.609	21	5	25

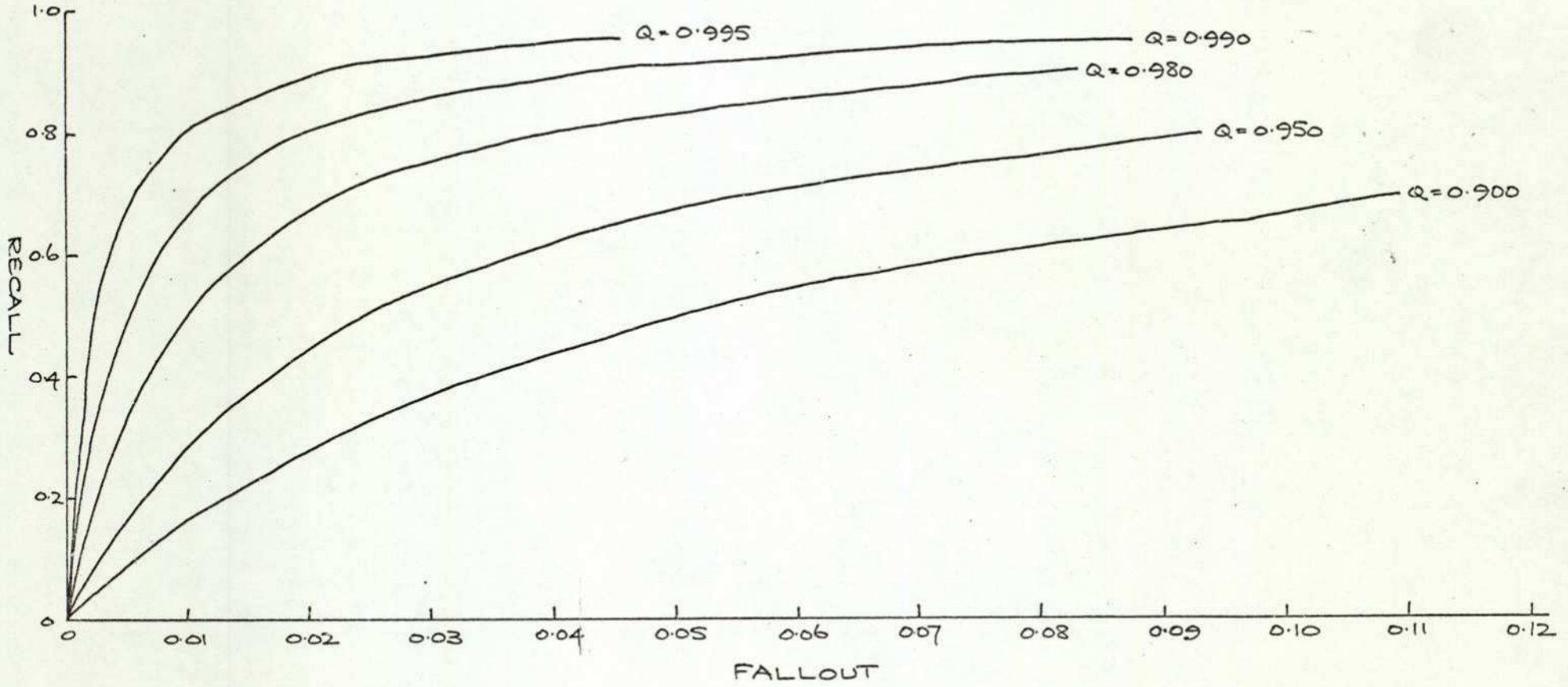


FIGURE 22 Recall - fallout diagram, showing lines of constant Q

In order to discern more easily whether or not the results do fall on Q values, it is preferable to plot the values in such a way as to give a straight line. Thus, if $Q = \frac{M - F}{M + F - 2MF}$, then $\frac{1}{M} = (\frac{1 - Q}{1 + Q})\frac{1}{F} + \frac{2Q}{1 + Q}$. A graph of $\frac{1}{M}$ against $\frac{1}{F}$ should give lines of gradient $\frac{1 - Q}{1 + Q}$ and intercept $\frac{2Q}{1 + Q}$. The disadvantage is that if the system is operating in any way efficiently Q will be high and thus the gradient will be low and the intercept approximately equal to one. The figures for one set (123) were plotted in this way (see Fig. 23), but straight lines could not be clearly distinguished.

The December results (set 123) were investigated in more detail in an attempt to see if there is another empirical relationship between M and F. The results were plotted as a M - log F graph in Fig. 24. Again no clear pattern could be seen. Linear relationships may be present, but it is not clear whether such relationships should have positive or negative gradients. Thus the existence of an exponential relationship ($F = e^{kM}$, where k is a constant) cannot be concluded.

A graph of log M against log F was also plotted (Fig. 25). Again no clear relationships can be discerned and doubt exists about the sign of the gradient. A negative gradient would imply a parabolic relationship, whereas a positive gradient would imply a hyperbola. Either of the latter types of relationship would be more acceptable than that of a negative exponential.

None of the possibilities mentioned would be a completely satisfactory explanation of the experimental facts. Consider the general case, as shown in Fig. 26.

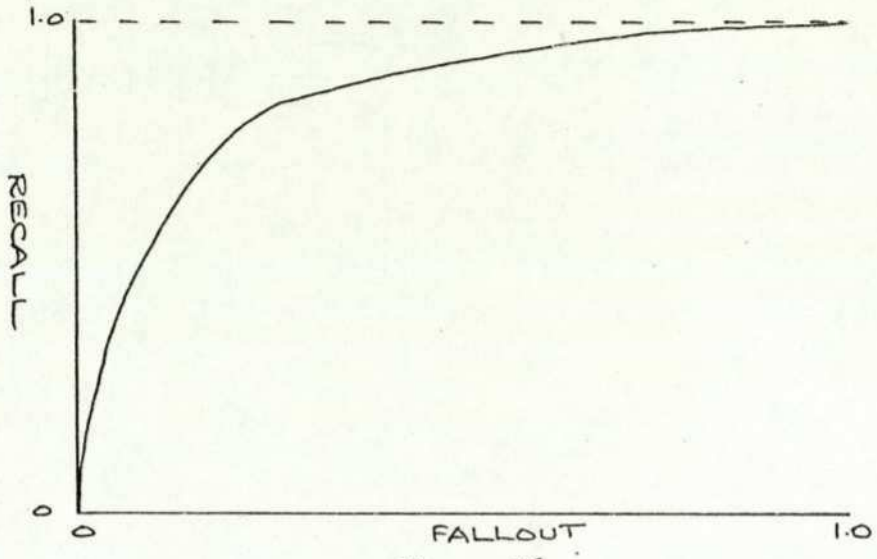


Figure 26

The line M = 1 must be the limiting value of the curve and therefore be an asymptote to other curves. From this point of view a parabolic curve would not be a satisfactory explanation, although of course the curve could be parabolic at the low values of F, which are generally the case in ISR system testing. At these low values it is reasonable that as the value M = 1 is approached, i.e. the last few of the relevant papers are selected, the number of irrelevant papers which are also selected by the system should increase. In view of the user's characteristics and the often apparently arbitrary way in which he differentiates between papers of fringe relevance, such a curve would be a realistic representation of the actual situation.

One of the difficulties of discerning the actual form of this relationship is due to the relatively few results available from this project. As the results for F tend to fall on a few discrete values, the trend is less clear. The number of uncontrolled variables may also be masking any operational characteristics.

3.2.8 The "generality" effect

Generality is a parameter which has been largely ignored in ISR system testing. Salton [86] discusses it in terms of small and large collections and quotes Robertson [80] as saying that as generality increases, precision increases and both recall and fallout remain constant. Such an assumption, if justified, implies that recall - fallout diagrams can be used to compare the performance of small and large systems and an adjusted precision can be calculated for transferring results from one system to another. Although the present research is not concerned with the comparison of two systems, it is nevertheless relevant to investigate the validity of extrapolating the present results to a similar system of a different magnitude.

Initially it is felt that there is no justification for assuming that recall and fallout are independent of generality. Considering the 2 x 2 contingency table, if C increases relative to N, i.e. G increases, it is reasonable to suppose that R also increases, but not necessarily in the same proportion as C, and therefore there is no reason to suppose that

recall is independent of generality. This is particularly applicable in small systems where a very small relative change in R or C can quite drastically affect the proportion or percentage. A similar situation is applicable to fallout, although as the quantity (N - C) is likely to be larger, the effect on the proportion, i.e. fallout, will be less apparent.

Relationships between recall, fallout and generality were investigated in a small-scale test of Metals Abstracts Index [39]. In this study tentative relationships of the form $M = -19.25\ G + 0.91$ and $F = 0.9\ G$ were found between the parameters.

In order to investigate the empirical aspects of these assumptions, the three parameters recall, precision and fallout were each studied in relation to generality.

Recall - generality

The recall - generality diagram shown in Fig. 27 indicated that the two parameters were not independent and that there was some sort of hyperbolic or exponential relationship present. Apparently there were four types of operation. The relationship was then further investigated by an approach similar to that used in Section 3.2.7 for recall - fallout. A graph of log M against log G for set 123 is shown in Fig 28. Although the points do not give a clear indication of a linear relationship, four (or five) straight lines could be tentatively drawn, of which one is the line $M = 1$. Such linear relationships would indicate

an equation of the form $M = k_1 + k_2 G^{k_3}$, where k_1, k_2 and k_3 are constants and $0 < k_3 < 1$.

As this graph is not completely convincing, an alternative approach was adopted and a graph of M against log G for the set was plotted in Fig. 29. Again it was possible to discern linear tendencies with a positive slope, although their existence could not be stated with conviction. Several different lines could be seen, where again one line would have the equation $M = 1$. Other linear relationships would suggest equations of the form $G = e^{kM}$, where k is a constant. A third approach was the graph of $\frac{1}{M}$ against $\frac{1}{G}$ shown in Fig. 30, where again linear relationships could be seen.

A modification of this in the form of a graph of M against $\frac{1}{G}$ is depicted in Fig. 31, where once more a linear relationship can be seen. Of the four representations, the latter is the most convincing with again four different lines being observable, but in this case the majority of the experimental results fall on these lines. The equation of one line is $M = 1$, and that of the other three $M = -k' \frac{1}{G} + k''$, where k' and k'' are constants. In each case $k'' = 1$ and k' is approximately 0.0021, 0.0042 and 0.0083 for the lines B, C and D respectively, i.e. $k'_D = 2k'_C = 4k'_B$. This would mean that the general expression of the relationship would be $G(1 - M) = k$. It should be pointed out that the generality range in this set of results is 0.004 to 0.10 approximately and thus one should not extrapolate this relationship to higher values of G.

Precision - generality

As has been previously discussed, a relationship between precision and generality is expected by, for example, Salton and Robertson. However, in Fig. 32 a relationship is not so obvious as it is between recall and generality. If one does exist, it appears to be also of a hyperbolic nature and consequently set 123 was investigated in greater detail. A similar procedure was adopted, i.e. graphs of log P - log G, P - log G, $\frac{1}{P} - \frac{1}{G}$ and $P - \frac{1}{G}$ were plotted and they are shown in Figs. 33 - 36. A number of straight lines can be seen in each case, but except in Fig. 36, the points do not fall into linear relationships in a particularly convincing way. The relationships in Fig. 36 were of the form $G(1 - P) = m$, where m is a constant. The values of m for the four straight lines were approximately 0, 0.0021, 0.0051 and 0.0125. In this case the constants increase by a factor of approximately 2.45.

Fallout - generality

The fallout - generality diagram (Fig. 37) is more difficult to interpret. Certainly there is a wide scatter of points. These could be interpreted as falling into five linear relationships of which the equations $F = 0$ and $F = 0.0021$ describe two. The other three relationships could be described by equations of the form $F = nG + q$, where n and q are constants. However, in view of the small number of points on each line, such suggestions are very tentative. The possible different linear

relationships did not coincide with the three modes of operation discussed in Section 3.2.9. As there was no indication of a hyperbolic/exponential relationship, the previous methodology was not pursued.

It does seem reasonable to conclude from these results that, within this generality range, recall and precision are not independent of generality. The recall - generality relationship appears to be hyperbolic, and the precision - generality relationship appears to have a similar type of expression.

3.2.9 Distribution of results

The distribution of results was studied in order to investigate more easily the way in which the system was operating. The average values of recall, precision and fallout given in Table 11 are arithmetic means. However, a frequency distribution and weighted means were also calculated for set 712.

As would be anticipated the differences between arithmetic and weighted means for the ratios are small, as shown in Table 14.

TABLE 14

Comparison of means (set 712)

	Recall	Precision	Fallout
Arithmetic	0.759	0.774	0.00368
Weighted	0.757	0.771	0.00379

The interval size selected for the determination of frequency distributions will affect the ultimate diagram. Initially interval sizes of 0.05 were selected for recall and precision distributions and the resulting graphs are shown in Fig. 38. Thus recall appears as a tri-modal function, with two sharp peaks (at 0.775 and 0.875) and one broad peak (at 0.575). Precision shows a different and quadri-modal distribution with a main peak (at 0.875) and three lesser peaks of decreasing strength at 0.750, 0.575 and 0.425.

By selecting 0.10 intervals, both recall and precision can be shown to have negative skew uni-modal distributions with peaks at 0.850 and long left hand tails from 0.400 to 0.650, as shown in Fig. 39.

A similar effect is observable with fallout figures. Intervals of 0.001 give a tri-modal distribution whose main modal value is at 0.00150. Intervals of 0.002 smooth out the lesser peaks in this distribution to give a skew (positive) uni-modal distribution, with the main peak at 0.0010 and a long right hand tail (Fig. 40).

A similar type of analysis was conducted for the December results in order to investigate more thoroughly the results for the three sets of primary parameters, i.e. sets 121, 122 and 123. Comparison of arithmetic and weighted means (Table 15) again shows that the results are closely similar, although in each case the weighted means give slightly worse figures than the

arithmetic means.

TABLE 15

Comparison of means (sets 121, 122 and 123)

	Arithmetic			Weighted		
	121	122	123	121	122	123
Recall	0.484	0.659	0.816	0.476	0.653	0.815
Precision	0.460	0.730	0.869	0.450	0.715	0.853
Fallout	0.01360	0.00619	0.00274	0.01404	0.00685	0.00308

For set 121 (Fig. 41) the recall and precision results both span the complete range 0 to 1, recall having three modes and precision four modes. Fallout also has a wide range, with two main modes and a long, undulating tail. Fig. 42 depicts the set 122 results where again recall and precision span the complete range, although in both cases the results have shifted towards the high end of the scale. The number of modes is retained in each case. For fallout the results have also shifted towards higher values, one effect of which is to shorten the length of the tail, while retaining the two main modes. The final set, 123, in Fig. 43, again demonstrate improved results for all three ratios. For recall the distribution is now essentially negative skew unimodal and the range of results is now decreased. For precision a similar effect has occurred except that the main mode consists of a more distinct combination of a sharp main peak with a broad shoulder. There is a second small mode at the end of the tail. For fallout

the two main peaks have combined into one, with a smoother right hand tail, whose length is again decreased.

For each set of recall and precision distributions intervals of 0.10 were used. The use of 0.05 intervals for set 123 again reveals a series of peaks (see Fig. 44). The interval sizes used for the fallout distributions were 0.0150, 0.0020 and 0.0010 for sets 121, 122 and 123 respectively because of the widely varying range of results.

The arithmetic means of the results which fall into the three main modes of operation for these sets are tabulated in Table 16. These figures are perhaps a better indication of the performance of the system than the overall arithmetic averages, although they give no indication of the number of results (= profiles) which fall into each category. Thus when comparing the results for sets 122 and 123 the conclusion might be drawn that the recall performance was worse in the latter case, whereas a greater proportion of the profiles fell into Mode I rather than Mode II and Mode II rather than Mode III, so that in fact the performance was improved.

The frequency distributions for the overall parameters, Q, (M x P) and G, were also studied for set 712 and are shown in Figs. 45, 46 and 47. The Q distribution is similar to that for fallout, whereas the distribution for (M x P) is more dispersed, as might be anticipated. The generality distribution is similar in shape to that for fallout and Q, but the main peak has a wider

TABLE 16

Arithmetic means for modes of operation

(sets 121, 122, 123 and 712)

	Mode III	Mode II	Mode I
Set 712			
Recall	0.589	0.771	0.892
Precision	0.498	0.753	0.899
Fallout	0.00177	0.00539	0.00818
Set 121			
Recall	0.087	0.436	0.732
Precision	0.241	0.690	0.991
Fallout	0.03746	0.01326	0.00338
Set 122			
Recall	0.321	0.640	0.878
Precision	0.338	0.705	0.972
Fallout	0.02067	0.00458	
Set 123			
Recall	0.200	0.443	0.876
Precision	0.500	0.738	0.967
Fallout	0.01735	0.00801	0.00113

spread and is less sharp. Q and G have tri-modal distributions, with one main peak and two lesser ones; (M x P) has a quadri-modal distribution with two main peaks and two lesser ones.

In so far as it is meaningful to use frequency distributions for the analysis of such a small number of results, these analyses indicate that in general there appears to be three (or four) main modes of operation of the system. If this is a valid conclusion, one would expect each mode of operation to produce results which fall into a normal distribution, a situation which is apparently present in these results. The elimination of some of the human variables has produced distributions of results which are more nearly normal and therefore are perhaps truer representations of the system and its errors. These distributions indicate that the use of scatter diagrams and averages does not reflect the true performance of a system which is not operating in a single mode. The factors determining these modes are discussed in Section 3.3.3.

3.2.10 The value of N

The value of N, the size of the total document collection, has been mentioned in Section 3.2.1. In the first City University relational indexing project [32, 36] this problem was discussed and two values of N were used in calculating the results. The first value was the size of the total document collection and the second the size of the collection of documents which contained one term in common with the question, that is to say, the document collection

equivalent to Cleverdon's one-term co-ordination level. In this first project no significant difference was found in the results and there was no correlation between performance and the second value of N.

In view of these results there was little to be gained in repeating the test. Hence, with the methodology adopted in the present project it seemed that there were only two possible values of N that could be realistically adopted. The first of these is the size of the total collection (= 2825). The second value is the size of the collection scanned by the users, as this is the set from which both C and L were selected. In the majority of cases this is also 2825 as most of the users scanned the five sections in making their selections. However, a small number of users scanned only some of these five sections and for these the value of N could be taken as the numerical size of the sections actually scanned.

An error may creep into the results at this stage. One vulnerable aspect is that although the user(s) may have claimed to be scanning all five sections, they might in fact have been scanning one or two of them. For a user whose interests were mainly contained within, for example, the structural hardening section, it would be very hard to tell whether or not he was scanning the section on chemical and electrochemical properties because the chance of a relevant paper occurring in the latter section is very small, and vice versa. Another possibility of what is essentially the same problem is the scanning of more sections than the five delineated for the project. If the user

was scanning other sections for the same or indeed wider interests, the resultant effect on his selections becomes an unknown quantity. For example, a user interested in chemical and electrochemical properties may well have been scanning the corrosion section and if in fact a high number of potentially relevant papers had occurred in the latter section, this might have had a restrictive effect on his selection of possibly relevant or fringe papers from the chemical and electrochemical section. Only in the cases of those relatively few users to whom photocopies of the appropriate five sections had to be sent, can it be reasonably safely assumed that they were unlikely to be affected by the latter influence.

The adopted definition of N affects the values for fallout and generality. In many systems the value of N is so large compared with the other three values of the 2 x 2 contingency table that it swamps the effects of changes in these values. However, a large change in N can have quite a significant effect on the calculated values of the ratios, without necessarily altering the real performance of the system. In this research it would be difficult to assign an exact interpretation on the change in the value of N as the number of users who were openly scanning less than the five sections was small.

The effects were investigated for set 123 and the changes in the recall - fallout diagram are shown in Fig. 14. Of the eight possible changes (i.e. users who were known to be scanning less than all five sections), five values of fallout equal to zero remain unchanged, so that only three of the points

are altered. However, for generality the changes are more numerous and the effects on the recall -, precision - and fallout - generality diagrams are shown in Figs. 48 and 49. In the recall - generality diagram the effect in several cases is to move the point from one operating curve to another. The effect is more noticeable in the fallout - generality diagram as in three instances both fallout and generality are modified. However, the number of results is inadequate for any general conclusions to be drawn.

3.2.11 An "overall percentage"

Since Cranfield II there has been some controversy over the methods used to obtain average values of recall, precision and the other parameters. Cleverdon and Keen [18] discuss the difference between the "average of numbers" and the "average of ratios", demonstrating that the average of ratios may give a better performance figure, especially for precision. They also mention Salton's suggestion that the average of numbers is document-oriented and the average of ratios is query-oriented. The average of ratios has been used in the present project for the majority of the results, including all those discussed to date. However, in a few instances the average of numbers has been used as this method has the advantage that it is considerably easier to calculate. Allowing for the inherent inaccuracies in the results, the difference between the figures calculated by the two methods is so small as to be insignificant.

Apart from these two methods of averaging, there is a third measure which is an assessment of the system as a whole. In this calculation L is the number of abstracts selected by the system for all users and is different from the ratio of numbers value as each unique (selected) abstract is counted only once. Similarly C is the number of unique abstracts selected by all the users from the total collection. R is the number of abstracts selected which are relevant to a profile, i.e. the number of coincidences between L and C. This calculation was carried out for set 123 and the averages are shown in Table 17 in comparison with the results obtained by the other two methods. As can be seen the differences between the average of numbers and the average of ratios are small, but the overall average gives results which are higher proportions than those obtained by the other methods, an indication that the system is working well at selecting items which are relevant to someone's profile. The averages are high also because there is a reasonable amount of duplication contained within each of the primary parameters, R, C and L.

TABLE 17

Comparison of averages (set 123)

	Average of ratios	Average of numbers	Overall average
Recall	0.816	0.801	0.896
Precision	0.869	0.873	0.922
Fallout	0.003	0.0025	0.083
Generality	0.023	0.021	0.521

Such an analysis does have the advantage that it enables some comparison between the performance of the separate sections. The averages for each section are tabulated in Table 18.

TABLE 18

Performance of separate sections of Metals Abstracts

Section	Recall	Precision	Fallout	Generality
1.2	0.840	0.895	0.136	0.579
1.4	0.929	1.000	0.000	0.757
3.2	0.895	0.810	0.190	0.475
3.3	0.911	0.962	0.286	0.444
3.4	0.957	0.978	0.200	0.485

The generality figures indicate that approximately half the abstracts in each section are relevant to some profile. In actual fact the amount of duplication was relatively limited, no abstract being selected by more than five users. Four abstracts were selected by five users, twelve by four, 29 by three; the remainder were selected by one or two users. It was thought that perhaps a relatively small subset of the data base might be selected by the users, with much greater overlap than that obtained, as has been suggested by work carried out at Case Western Reserve University [45]. The present results are not altogether unexpected in view of the wide range of interests and comparatively small numbers of users. Whether or not a larger number of users would produce similar generality figures is a matter for conjecture.

With regard to the individual sections, Table 18 indicates that section 1.4 gave the best results and also had the highest generality, although section 3.4 also gave good results with lower generality. This may be because of the more specific nature of their contents, i.e. "structural hardening" and "chemical and electrochemical properties" respectively.

3.2.12 The 0/0 situation

The problem of measures which have values of 0/0 is one which occurs when calculating ratios and averages of ratios of individual results. In calculating the average of numbers the problem does not arise unless all the results are zero - a highly improbable situation. The possibility of a 0/0 result occurs mainly when calculating recall and precision. There are three main situations, as follows:

(i) $R = 0$, $L = 0$, $C = 0$, therefore $\frac{R}{C} = \frac{0}{0}$; $\frac{R}{L} = \frac{0}{0}$.

This can be regarded as a good result, in so far as there were no relevant documents in the collection and the system did not select any irrelevant documents. In this case recall and precision might both be equated to 1.0.

(ii) $R = 0$, $C = 0$, $L = x$, therefore $\frac{R}{C} = \frac{0}{0}$; $\frac{R}{L} = 0$.

This is a bad result as the system is selecting irrelevant documents. The extent of the "badness" depends largely on

the actual value of x. For example, if L = x = 1, the user is presented with one document which is irrelevant. If however, x = 200, then the system is not performing at all well. However, the ratios as such make no distinction between these two cases.

(iii) R = 0, C = x, L = 0, therefore $\frac{R}{C} = 0$; $\frac{R}{L} = \frac{0}{0}$.

Again this is a bad result as the system is missing relevant papers, and the degree is dependent on the actual value of x.

It can be seen from Table 19 that the number of instances of these three cases in each set is small and the majority of these are of the first type, indicating good results. There is little difference between XY1, XY2 and XY3 sets for each month, but there are more occurrences with the relevance assessments separated than with the combined result, as would be expected. The number of instances falls off with the cumulative results, there being only one case in set 712.

3.2.13 Categories of relevance

It was stated in Section 2.6 (Matching) that the users were asked to assign the numbers of relevant abstracts to two categories, namely "Abstracts of definite interest" or "Abstracts of possible interest". The former category was termed relevance 1 and the latter relevance 2. The results discussed so far have considered the combined relevance (1 + 2).

TABLE 19

Instances of 0/0 ratios

Set	$\frac{R}{C} = \frac{0}{0}$; $\frac{R}{L} = \frac{0}{0}$	$\frac{R}{C} = \frac{0}{0}$; $\frac{R}{L} = 0$	$\frac{R}{C} = 0$; $\frac{R}{L} = \frac{0}{0}$	Total
071	1	1	-	2
072	-	-	-	-
073	2	-	-	2
081	4	3	1	8
082	4	-	1	5
083	4	-	2	6
091	4	-	2	6
092	4	-	2	6
093	5	-	1	6
101	2	3	-	5
102	2	1	-	3
103	3	-	-	3
111	1	-	-	1
112	1	-	-	1
113	1	-	-	1
121	3	1	-	4
122	3	-	-	3
123	3	-	-	3
123 (rel 1)	6	3	7	16
123 (rel 2)	3	2	1	6
712	1	-	-	1

The allocation of abstracts to relevance categories is highly subjective and depends on the user's own interpretation of "definite interest" and "possible interest". Also there is no quantitative method for ranking the output from a relationally indexed system. In this project an attempt was made to rank documents according to the following procedure:

1. Each document was assigned an initial value of 10.
2. One point was subtracted from each separate string that was additional to the match between the document and the profile.
3. One tenth of a point was deducted for each additional relation in the structure.
4. The remaining value was the number assigned to the document.

For example, for a profile which consisted of

Cu /(electrical conductivity

a document would be selected which had the following analysis:

/; heating
Cu /(electrical conductivity /- measuring
/+
crystal /(structure /; studying /; X-ray diffraction.

Thus the value of this document in relation to the profile given would be $10 - (2 \times 1) - (6 \times 0{\cdot}1) = 7{\cdot}4$.

It was hoped that these values would give a means of ranking the output, the documents with a value of 10 representing

complete matches and therefore being the most relevant. In practice it was found that the numbers assigned in this way did not bear any relationship to the relevance assessments given by users. Thus, although it might be anticipated that the presence of additional strings and the number of additional relations would detract from the relevance of an abstract, it appears that perhaps the actual concepts themselves are of greater significance.

As this quantitative assessment failed, a qualitative method was used in which the system operator assigned a value of 1 or 2 to the relevance of each selected abstract. It was these values which were used for the comparison of relevance assessments. Thus results were calculated for the parameters R, C and L on the basis of a 2 x 2 contingency table for relevance 1 and also relevance 2. Recall - precision diagrams are plotted for relevance 1 and relevance 2 for set 123 in Fig. 50 and similar recall - fallout diagrams are shown in Fig. 51. From these diagrams it can be seen that the results for the separate relevance categories were very variable and also that there were many instances of 0/0. Because the relevance categories are treated independently rather than as a 3 x 3 contingency table, many of the results where the relevance assessment did not agree are not included in either set of the final figures, appearing only in combined relevance calculations.

The averages (of ratios) are given in Table 20, from which it can be seen that the combined relevance results are better than either of the individual sets. Also the system is better at predicting possibly relevant documents than definitely relevant ones.

TABLE 20

Average results for relevance categories (set 123)

	Recall	Precision	Fallout
Relevance 1	0.2989	0.4553	0.00333
Relevance 2	0.5626	0.5579	0.00597
Relevance (1 + 2)	0.8159	0.8689	0.00274

3.2.14 Size of data base for testing

As mentioned in Section 2.2 (Data base), information on the size of data base needed for adequately testing an ISR system is not known. In the light of the results from the first City University project, a statistical analysis [9] indicated that it would have been necessary to index only 64 papers, although the appropriate number of questions was not obtainable. This viewpoint was received with a certain amount of incredulity and it was hoped that this project would provide some facilities for exploring this area, as a relatively large data base was used.

The values for N obtained by cumulating are as follows:

Cumulation	N
July	723
July + August	1171
July to September	1460
July to October	1758
July to November	2345
July to December	2825

If the results had been more consistent over the duration of the project, the effect of increasing the size of the data base could have been investigated. However, as the results improved with time, it was inevitable that the cumulative results should also improve as the size of the data base increased.

In the future it is hoped to conduct a separate experiment using this data base and these results to sample in cumulative randome groups of perhaps 250, and thus to see if there is a minimum value, less than 2825, which appears necessary in order to obtain reasonably reliable results.

3.2.15 Number of questions required

The number of questions or profiles necessary to test a system adequately is another area of testing methodology where existing information is scant. There are various aspects to the problem, one of which is the average of ratios. A relatively small number of questions are necessary to provide an average, the exact number depending on the range and scatter of the results and the degree of accuracy required. For example, the average value of recall and precision against number of results for sets 121, 122 and 123 are shown in Figs. 52 and 53. These show that the variation in the average is rapidly damped and relatively small after about 20 results. To some extent the degree of damping depends on the order in which the results are averaged. In this case they were in the order in which the users volunteered. Ideally the order should be random, but this order could be regarded

as reasonably unbiassed, and indeed to an extent random. The effectiveness of this type of analysis depends on the effect of an additional question and obviously the greater the number of questions the lesser the impact an additional one will have, even if the value is widely removed from the average. For example, if the average of 100 questions is 50.0% and another question is posed which has a result of 0%, the new average will be 49.504%, i.e. irrespective of its value, the change in the average value will be so small as to be insignificant.

Formalising this into a general case,

let x_i = the value of the i^{th} result, $0 \leq x \leq 1$,

then the average for n questions, $A_n = \frac{\sum_{i=1}^{n} x_i}{n}$

For (n + 1) questions, $A_{n+1} = \frac{\sum_{i=1}^{n+1} x_i}{n+1} = \frac{nA_n + x_{n+1}}{n + 1}$

Thus the difference between A_{n+1} and A_n is

$$A_{n+1} - A_n = \frac{nA_n + x_{n+1}}{n + 1} - A_n$$

$$= \frac{nA_n + x_{n+1} - nA_n - A_n}{n + 1}$$

$$= \frac{x_{n+1} - A_n}{n + 1}$$

Suppose that this difference should be less than 1%, then

$$\pm \frac{1}{100} A_n = \frac{x_{n+1} - A_n}{n + 1}$$

In order that this equation can be solved for n, it is necessary to express x_{n+1} in terms of A_n. For example, suppose that $x_{n+1} = 2A_n$, then

$$\pm \frac{1}{100} A_n = \frac{2A_n - A_n}{n + 1}$$

$$\frac{1}{100} = \frac{1}{n + 1}$$

$$n = 99$$

In the present system, $A_n = 0.65$ for recall in set 122. Assume that $x_{n+1} = 1$, i.e. $x_{n+1} = 1\frac{1}{2} A_n$ and a 1% difference is required, then the number of questions required would be

$$\frac{1}{100} A_n = \frac{1\frac{1}{2} A_n - A_n}{n + 1}$$

$$\frac{1}{100} A_n = \frac{A_n}{2(n + 1)}$$

$$n = 49.$$

For a 5% difference,

$$\frac{1}{20} A_n = \frac{1\frac{1}{2} A_n - A_n}{n + 1}$$

$$\frac{1}{20} = \frac{1}{2(n + 1)}$$

$$n = 9$$

With a calculation of this type it can be seen that the number of questions needed depends on the type of results being achieved by a system.

However, this takes no account of the other facets of the system. If, as in the present project, the parameters show a multi-modal distribution, the average value of any one parameter is not as meaningful as it might be.

3.2.16 Error analysis

As previously mentioned in Section 3.1.1, a certain number of errors were considered "acceptable" in so far as they were human errors not associated with the operation of the system. There were of course others which did not relate to the operation of the system and an error analysis was conducted.

The allocation of errors to categories is in itself a subjective operation as it is often extremely difficult to define the exact cause of the failure, if indeed it is attributable to only one factor and not to a combination of factors. There are also very many human variables which cannot be measured with our existing knowledge of ISR system testing and in fact it is very doubtful whether all of them have yet been enumerated or recognised, although attempts have been made by various people (e.g. Cuadra and Katter [22], Saracevic and Rees [88]).

In this research a total of 27 error types were categorised, of which a few occurred rarely if at all. They were grouped into "acceptable" and "non-acceptable" and "user" and "system". All the user errors were regarded as "acceptable". The "system" in this context is taken to include both the relational

indexing aspects and the human errors associated with the searcher. The error types are defined as:

User errors - acceptable

1. User selection error: the omission by the user of relevant items or the inclusion of irrelevant items.

2. User profile error: an error in the profile caused by the user not including an aspect of interest, including aspects in which he is not interested, or changing his interests without reporting them to the system.

3. Sixth sense: an item included by the user as of interest, although it has no possible relation to the profile as stated. This could be regarded as an extreme form of user profile error.

4. Language error: the item is rejected because the language is not read by the user, even though the item may be relevant.

5. User's idiosyncrasies: apparently inexplicable reasons why the user accepted/rejected a seemingly irrelevant/relevant item.

System errors - acceptable

6. Selecting error: the non-selection of an item from the data base caused by the selector missing it, turning over two cards together or otherwise failing to retrieve it for no apparent reason. Also the selection of a non-relevant item for no

apparent reason.

7. Typing error: an error in heading the card, a pure typing mistake or miscopying of the original handwritten analysis. Misfiling is included in this error type.

8. Technical knowledge: lack of technical knowledge by the selector meant that a relation between the profile and indexed abstract was not observed. This error type was mainly due to lack of full comprehension of technical terminology and is distinct from the technical implication error.

System errors - non-acceptable

9. Indexing error/word: an error in the relational indexing of the item brought about by the use of the wrong word or concept.

10. Indexing error/relation: the wrong relation used in the indexing of an item.

11. Indexing error/depth: the item was indexed in too great depth, causing selection when not relevant to the profile, or, less frequently, indexing in insufficient depth for selection of a relevant item.

12. Profile interpretation: a technical error in interpreting the user's profile.

13. Profile construction: a technical error in constructing the relationally indexed version of the profile.

14. Technical implication: this error can be caused by an item

being seen to be relevant to a profile, although there does not seem to be a direct route from the profile to the abstract by any of the searching strategies, e.g. conceptual browsing, generic browsing, condensations, i.e. the sum of the concepts and relations approximately add up to the profile. Alternatively the error may be brought about by one or two of the concepts taken together implying other concepts which are present in the profile.

15. Generic browsing/omission: the item could have been retrieved by generic browsing.

16. Conceptual browsing/omission: the item could have been retrieved by conceptual browsing.

17. Condensation/omission: the item could have been retrieved by one or more condensations.

18. Related browsing/omission: the item could have been retrieved by related browsing.

19. Lower co-ordination: the item could have been retrieved by using a lower co-ordination level search strategy.

20. Generic browsing/incorrect: the item was selected because of incorrect generic browsing.

21. Conceptual browsing/incorrect: the item was selected because of incorrect conceptual browsing.

22. Condensation/incorrect: the item was selected because of an incorrect condensation.

23. Too theoretical/academic: the item was considered either too theoretical or too academic by the user, although it apparently matched his profile (usually by industrial users).

24. Too practical/applied: the item was considered too practical or too applied by the user although it apparently matched his profile (usually by academic users).

25. Too general: the item was considered to be of too general an application to a user's specific requirement. Such errors were usually caused by generic browsing.

26. Too specific: the item was considered to be too specific or too detailed for the user's more general interests. Such errors were caused mainly by either generic browsing down the classification or by the match plus strategy where the additional concepts and relations showed that the item had specific constraints that made it too detailed an aspect of the general problem.

27. Unspecified: the error either did not fall into any of the other categories or, more frequently, the user did not specify why the item was relevant/irrelevant and no reason could be deduced by comparing the item and the profile.

Each error was analysed (for each user for each of the six issues) and assigned to the most suitable of the above categories. In some profiles the same type of error tended to occur throughout the six months; in other profiles the errors varied from month to month. A summary of the errors by months, with the totals for each

type is shown in Table 21.

This analysis shows that the largest single cause of error is user selection (25.6%) and that user profile (17.6%) and system selection (10.2%) are the next highest (a combined total of 53.4%). Acceptable errors accounted for 63.8% of the overall error occurrences. Of the non-acceptable errors, technical implication (6.5%) and unspecified (6.2%) are the two largest groups. The other errors that appear in significant proportions are sixth sense, indexing/depth, profile interpretation, generic browsing/ omission, conceptual browsing/omission and too theoretical/academic.

The number of errors per month will obviously bear some relationship to the number of abstracts and in fact there is a very close correlation. Presumably it is coincidence that the total number of errors is equal to the total number of abstracts, especially in view of the fact that only about half the abstracts were utilized. The number of errors per user also varies considerably and this may depend to some extent on the number of abstracts selected. However, a number of other factors are also involved, such as how consistent the user is in his selection, how precise is his original profile, etc. Some of the errors follow the size of the monthly collection, e.g. selecting, user selection, etc. Others fall off with time as the system more nearly approaches an equilibrium situation. For example, user profile and profile interpretation had their largest values in the July issue as might be expected and thereafter had a nearly constant value. Profile construction errors were eliminated after the first two months.

TABLE 21

Summary of error occurrences

	Error type	7007	7008	7009	7010	7011	7012	Total	%
	User - acceptable								
1.	User selection	137	76	85	89	186	149	722	25.6
2.	User profile	223	56	49	55	75	40	498	17.6
3.	Sixth sense	32	14	8	8	15	19	96	3.4
4.	Language	2	5	4	1	3	5	20	0.7
5.	User's idiosyncrasies	-	9	1	4	9	9	32	1.1
	(Sub-total)	394	160	147	157	288	222	1368	48.4
	System - acceptable								
6.	Selecting	86	68	25	14	53	43	289	10.2
7.	Typing	2	1	1	1	2	5	12	0.4
8.	Technical knowledge	66	22	10	7	12	16	133	4.7
	(Sub-total)	154	91	36	22	67	64	434	15.4
	(Sub-total - acceptable)	548	251	183	179	355	286	1802	63.8
	System - non-acceptable								
9.	Indexing/word	8	15	4	4	6	12	49	1.7
10.	Indexing/relation	-	-	1	-	2	-	3	0.1
11.	Indexing/depth	26	18	8	9	16	15	92	3.3
12.	Profile interpretation	46	7	14	4	7	4	82	2.9
13.	Profile construction	3	2	-	-	-	-	5	0.2
14.	Technical implication	44	39	15	19	36	30	183	6.5
15.	Gen. browsing/omission	29	24	11	15	17	26	122	4.3
16.	Con. browsing/omission	36	30	7	17	20	10	120	4.3
17.	Condensation/omission	4	2	3	1	6	7	23	0.8
18.	Rel. browsing/omission	2	2	2	3	5	11	25	0.9
19.	Lower co-ordination	2	-	1	4	4	3	14	0.5
20.	Gen. browsing/incorrect	3	-	1	-	3	-	7	0.3
21.	Con. browsing/incorrect	1	-	-	-	-	-	1	0.0
22.	Condensation/incorrect	1	-	-	-	-	-	1	0.0
23.	Too theoretical/academic	22	18	6	8	22	7	83	2.9
24.	Too practical/applied	9	-	-	-	6	1	16	0.6
25.	Too general	1	-	-	-	1	1	3	0.1
26.	Too specific	-	-	-	3	16	1	20	0.7
27.	Unspecified	48	36	19	13	36	22	174	6.2
	(Sub-total)	285	193	92	100	203	150	1023	36.2
	TOTAL	833	444	275	279	558	436	2825	100.0

It is inevitable that the error analysis for this project should be very dependent upon this system, as is the case with other evaluations, for example, UKCIS [7] and INSPEC [2] evaluations. Thus all the errors could be classed as human errors, inasmuch as they contain some human element, and again this is inevitable with an essentially manual system.

3.2.17 Effects of searching strategies

A complete analysis of the effects of searching strategies was performed in respect of the July and December issues and the cumulative total, together with a partial analysis for all issues. In the complete analysis recall and precision ratios were calculated for each strategy individually and for combinations of match plus with the other strategies. These results are tabulated in Table 22. The partial analysis consisted of the results of cumulating all strategies in their order of effectiveness and also of match plus, conceptual browsing and condensations (logical jumps) and these results are depicted in Table 23.

The analysis shown in Table 22 indicated that in general the individual strategies made small contributions to the total recall and that the precision of the various strategies varied from month to month. Overall, the match plus strategy is the most precise, but the monthly variation is further elaborated in Table 24 which shows the precision of each strategy for each issue. Of the

TABLE 22

Effects of searching strategies

Strategy	Set 072		Set 122		Total	
	M	P	M	P	M	P
Match plus (MP)	0.238	0.455	0.425	0.741	0.351	0.664
Generic browsing (GB)	0.171	0.527	0.190	0.791	0.175	0.636
Conceptual browsing (CB)	0.097	0.609	0.075	0.702	0.075	0.591
Condensations (LJ)	0.061	0.432	0.055	0.727	0.064	0.500
Related browsing (RB)	0.016	0.250	0.011	0.417	0.016	0.442
MP + GB	0.409	0.483	0.614	0.756	0.526	0.654
MP + CB	0.335	0.491	0.500	0.735	0.426	0.650
MP + LJ	0.300	0.451	0.480	0.739	0.415	0.632
MP + RB	0.254	0.434	0.436	0.726	0.367	0.649

TABLE 23

Partial analysis of effects of searching strategies for each month

	July		August		September		October		November		December		Total	
Strategy	M	P	M	P	M	P	M	P	M	P	M	P	M	P
Match plus	0.238	0.455	0.263	0.570	0.307	0.744	0.393	0.815	0.393	0.770	0.425	0.741	0.351	0.664
MP + GB	0.409	0.483	0.349	0.559	0.544	0.758	0.522	0.765	0.572	0.717	0.614	0.756	0.526	0.654
MP + GB + CB	0.506	0.503	0.405	0.562	0.608	0.736	0.584	0.731	0.630	0.692	0.690	0.749	0.600	0.645
MP + GB + CB + LJ	0.566	0.494	0.459	0.531	0.694	0.715	0.635	0.705	0.692	0.668	0.744	0.748	0.664	0.628
Total	0.582	0.481	0.479	0.533	0.707	0.707	0.645	0.702	0.712	0.671	0.756	0.739	0.680	0.622
MP	0.238	0.455	0.263	0.570	0.307	0.744	0.393	0.815	0.393	0.770	0.425	0.741	0.351	0.664
MP + CB	0.335	0.491	0.319	0.573	0.372	0.713	0.454	0.760	0.451	0.724	0.500	0.735	0.426	0.650
MP + CB + LJ	0.396	0.481	0.373	0.535	0.457	0.689	0.506	0.723	0.514	0.686	0.555	0.734	0.490	0.625

TABLE 24

Precision of various strategies

Strategy	7007	7008	7009	7010	7011	7012	Total
Match plus	0.455	0.570	0.744	0.816	0.770	0.741	0.664
Generic browsing	0.527	0.525	0.775	0.644	0.623	0.791	0.636
Conceptual browsing	0.609	0.583	0.594	0.529	0.517	0.702	0.591
Logical Jumps	0.432	0.385	0.595	0.500	0.500	0.727	0.500
Related browsing	0.250	0.636	0.444	0.600	0.769	0.417	0.442

pairs of strategies, match plus with generic browsing gave the highest recall and also the highest precision overall, apart from the July issue when the best result was obtained by match plus with conceptual browsing.

The results for various combinations of strategies can be plotted on recall - precision diagrams, as shown in Fig. 54. By plotting the combinations in a cumulative manner, it can be seen that the various strategies improve recall, sometimes with a slight improvement in precision, sometimes with a slight deterioration in precision. The lines are virtually vertical.

However, such a procedure could be open to the criticism levied at Cleverdon's lines joining points obtained with different levels of co-ordination, in so far as generic and related browsing could be considered as posing a different question to the system. In reality this depends on the interpretation placed on the application of the browsing techniques. Conceptual browsing and condensations are not subject to such criticism as they are different ways of asking the same question. Thus these two techniques alone are combined with match plus and the resulting lines are shown in the lower diagram. The variation is more marked. The addition of condensations usually results in improved recall and a slight fall off in precision, whereas the combination of match plus with conceptual browsing improves recall and has a variable effect on precision. Thus in some months the lines are virtually vertical, whereas in others (notably October and November) they look more like the levels of co-ordination lines of Cleverdon.

These results give some indication of the way in which a system might be adapted to meet special requirements. For example, if high recall was of paramount importance then obviously one would use all possible selecting strategies. On the other hand, if high precision was the major criterion, then conceptual browsing alone should be used according to the July results, generic browsing alone according to the December results and match plus alone according to the overall results. A combination of this knowledge, together with a knowledge of which strategy is the most effective for an individual profile, obtainable from the data summarised in Table 10 (Section 3.1.11) indicates the way in which searching could be matched with the requirements of individual profiles. Such conclusions should however be treated with considerable care. A strategy which selects relevant items with the greatest precision is not necessarily a strategy which selects from the collection the items with the highest relevance.

3.2.18 Consistency of indexing

One of the more unexpected results from the first City University relational indexing project was that when the searches were made without including the relations, i.e. using the terms as keywords, the results were unchanged in 98 of the 127 questions [32]. The explanation suggested for this result was that the stricter control of relational indexing and concept organization (compared with keyword systems) induced the selection of better terminology which was effective without the relations. This implies that

control should also be exerted upon different indexers and hence inter-indexer consistency might be improved with this technique.

Most of the previous studies of inter-indexer consistency compare the number of terms assigned to a document by both indexers to the total number of terms assigned. Zunde and Dexter [102] use their concept of "fuzzy sets" to differentiate between the more important and less important terms which are assigned by two indexers in common, and this in general results in higher values for the consistency. This approach is extended to define quality and exhaustivity of indexing and they show that the amount of information indexed increases with number of indexers in a way analogous to the Bradford law of scattering. The relationship between inter-indexer consistency and the quality of indexing and the retrieval effectiveness is discussed by Cooper [20] who demonstrates that there is an (unknown) mathematical relationship between inter-indexer consistency and retrieval effectiveness. He is insistent that until such a relationship is defined, inter-indexer consistency should not be used as a measure of indexing quality.

The facilities for studying indexing consistency in the present project were somewhat limited, both by time and personnel. However, the previous results make it an interesting aspect of relational indexing, which, to be studied in depth, would need a larger research project. However, some comparative figures can be obtained. Although the majority of the relational indexing

was performed by the author, Farradane assisted with some sections. The largest amount indexed by Farradane was 184 of the 448 abstracts in the August issue, and hence that month (set 083) was used as the basis for the study. Because of the amount of time available and the amount of work to be done, it would have been unrealistic for both indexers to index the same material and thus facilitate a comparison of a type similar to that proposed by Zunde and Dexter. Instead a comparison is achieved by comparing recall, precision and fallout figures for the two subsets of the August data base. The assumption implicit in this approach is that if there was a marked inconsistency between the two indexers then the performance ratios would deteriorate, especially as the profiles were all indexed by the author. Similarly the percentage of non-acceptable errors in the total number of errors should differ if there was a significant inconsistency. The results of these comparisons, calculated by the ratio of averages method, are shown in Table 25 and from this it can be seen that there are no significant differences for retrieval from the two subsets of the data base.

A second aspect of the indexing was also considered by a similar methodology. The abstracts from the August issue indexed by the researcher were assigned a value of 1, 2, 3 or 4, which represented a subjective opinion by the indexer of the "difficulty" of the abstract when indexing. The value 1 was assigned to the "easiest" abstracts and the value 4 to the "most difficult". It is perhaps inevitable with a subjective evaluation of this type that the majority of the abstracts fell into categories 2 and 3 and it was for this reason that a 3- (or other odd number)

TABLE 25

Comparison of retrieval effectiveness for different indexers

	Indexers	
	PAYM	JELF
Recall	0.694	0.675
Precision	0.711	0.752
Fallout	0.00875	0.00368
Non-acceptable errors / Total errors (%)		
Selected non-relevant (L - R)	56.5%	59.6%
Relevant not selected (C - R)	78.6%	70.7%

valued scale was avoided. The objective of the analysis was to ascertain whether the "difficulty" of the abstract resulted in poorer quality indexing as reflected by less efficient retrieval. The mental indexing processes that occur when confronted with a "difficult" abstract cannot be analysed objectively. Obviously one has to think about the subject content and its meaning for a longer period of time and thus endeavour to assign the most accurate interpretation, in terms of concepts and relations, with which to represent it. Whether this longer period of thought results in a representation which is ultimately as "good" as that for easier abstracts, or whether the abstract is too difficult for full understanding and interpretation of its meaning to be

found, is questionable.

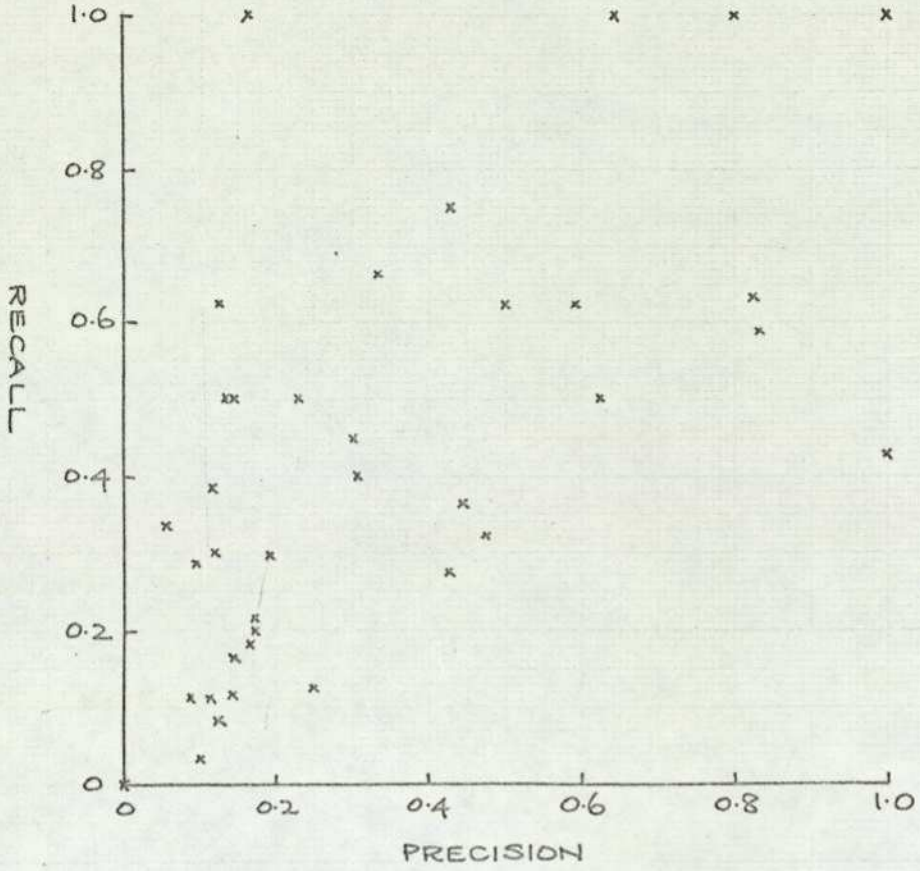
The analysis of the retrieval effectiveness is shown in Table 26 and from these results it appears that the effectiveness falls off with increasing difficulty of the abstract. Overall recall and precision decrease and fallout increases with difficulty. The percentage of non-acceptable errors in the (C - R) set increases with the degree of difficulty, although there is little correlation with the proportion of non-acceptable errors in the (L - R) set. Overall there is apparently a relationship between abstracts that the indexer considers difficult and the effectiveness of retrieval.

TABLE 26

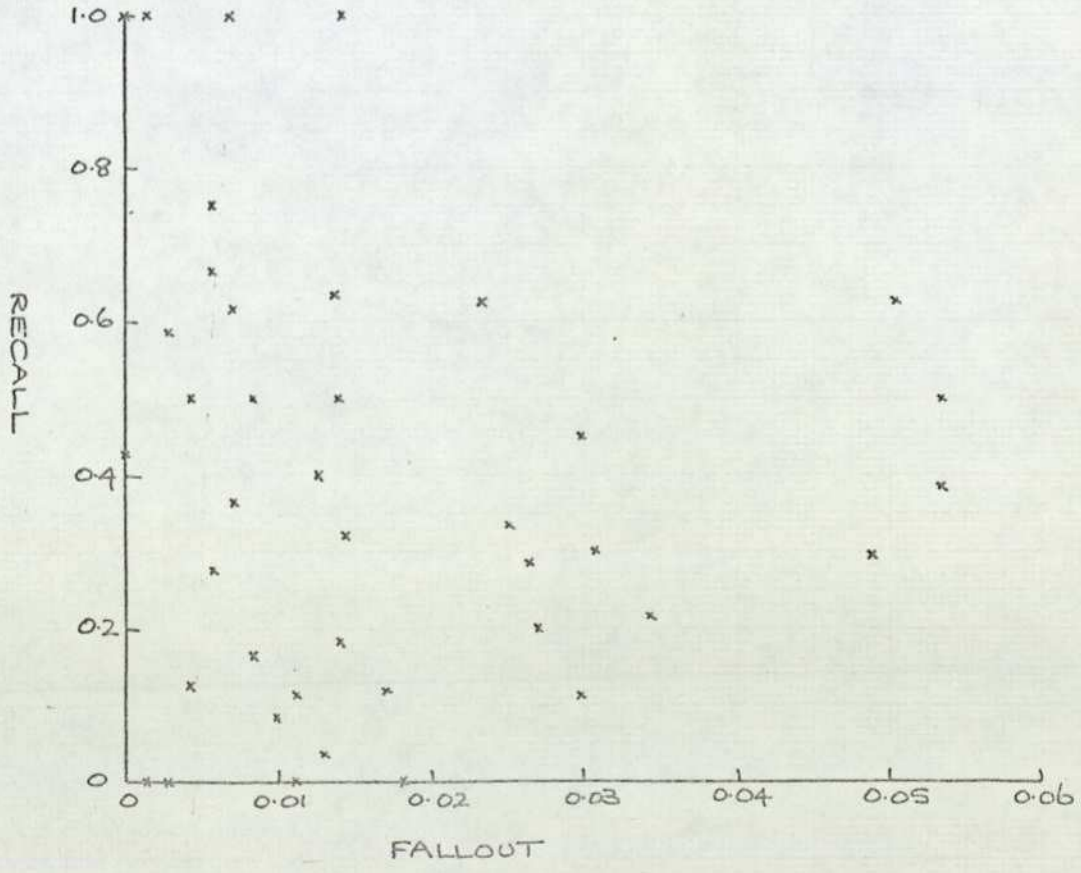
Comparison of retrieval effectiveness according to the "difficulty" of the abstracts

	Degree of "difficulty"				
	1	2	3	4	Unspecified
No. of abstracts in each category	4	85	145	18	12
Recall	1.000	0.790	0.698	0.304	0.800
Precision	1.000	0.771	0.667	0.700	0.667
Fallout	0.00000	0.00545	0.00618	0.00787	0.00400
Non-acceptable errors / Total errors (%)					
Selected non-relevant (L - R)	-	63.3	53.6	60.0	50.0
Relevant not selected (C - R)	-	68.0	78.0	94.1	100.0

Graphs relating to Section 3.2

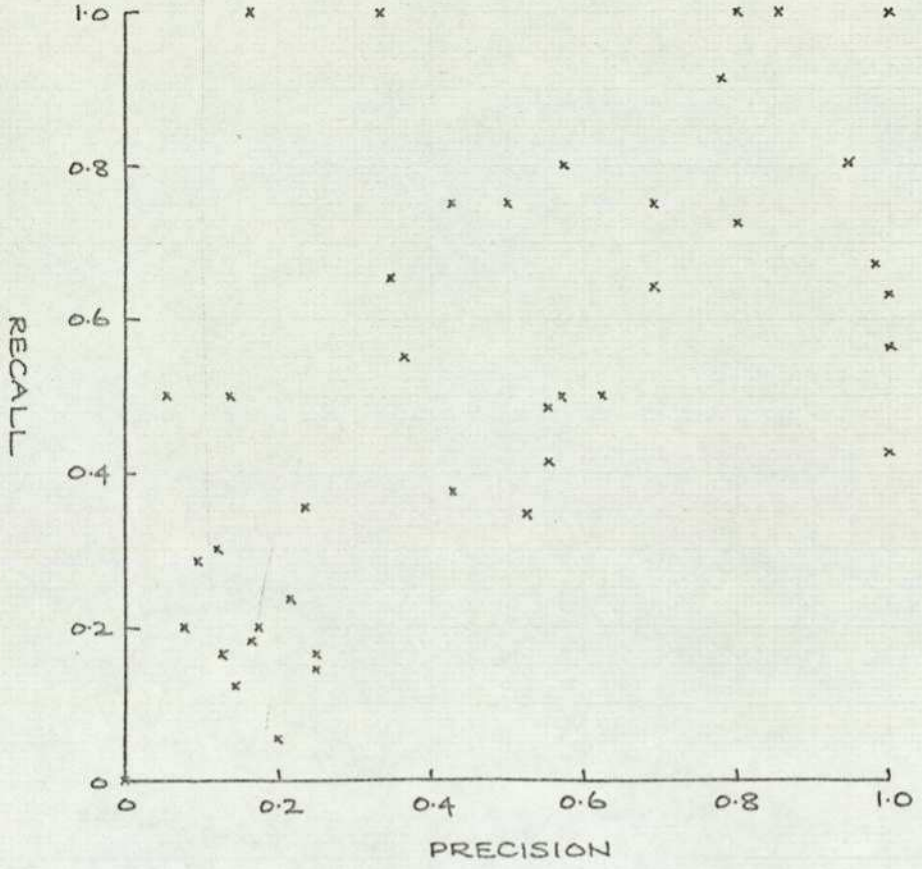


Recall - precision diagram (set 071)

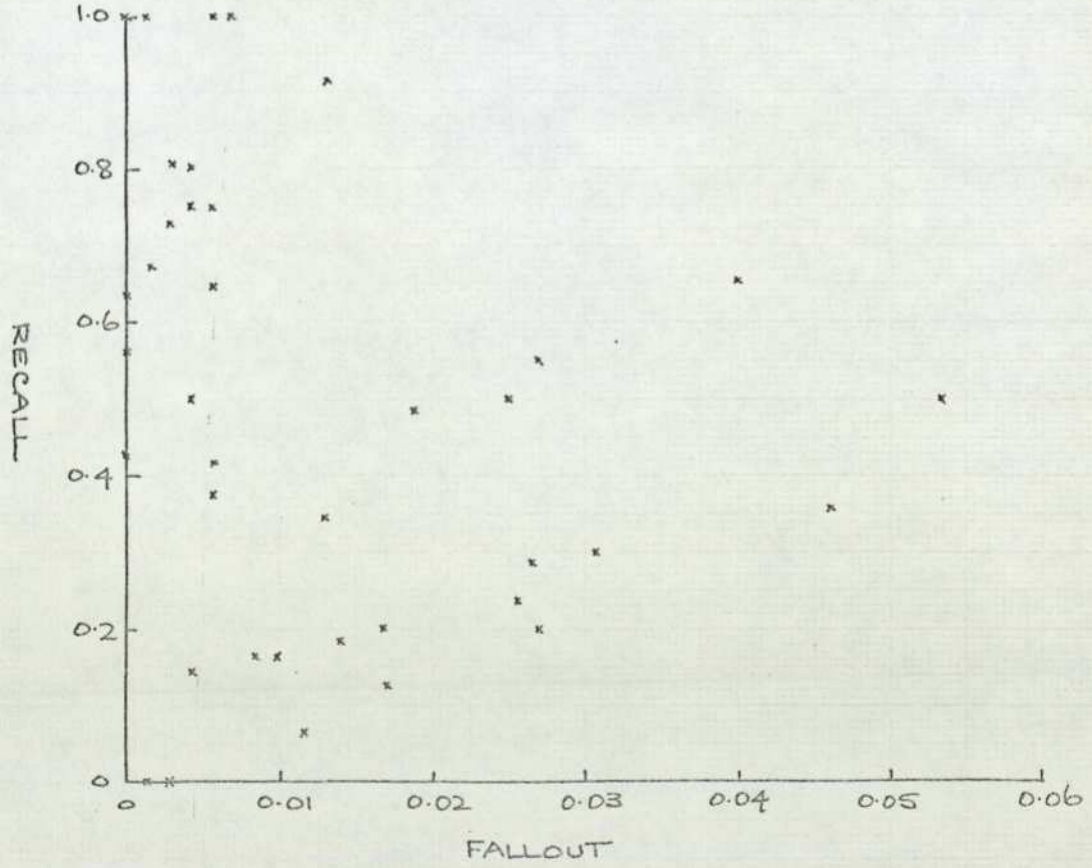


Recall - fallout diagram (set 071)

FIGURE 7

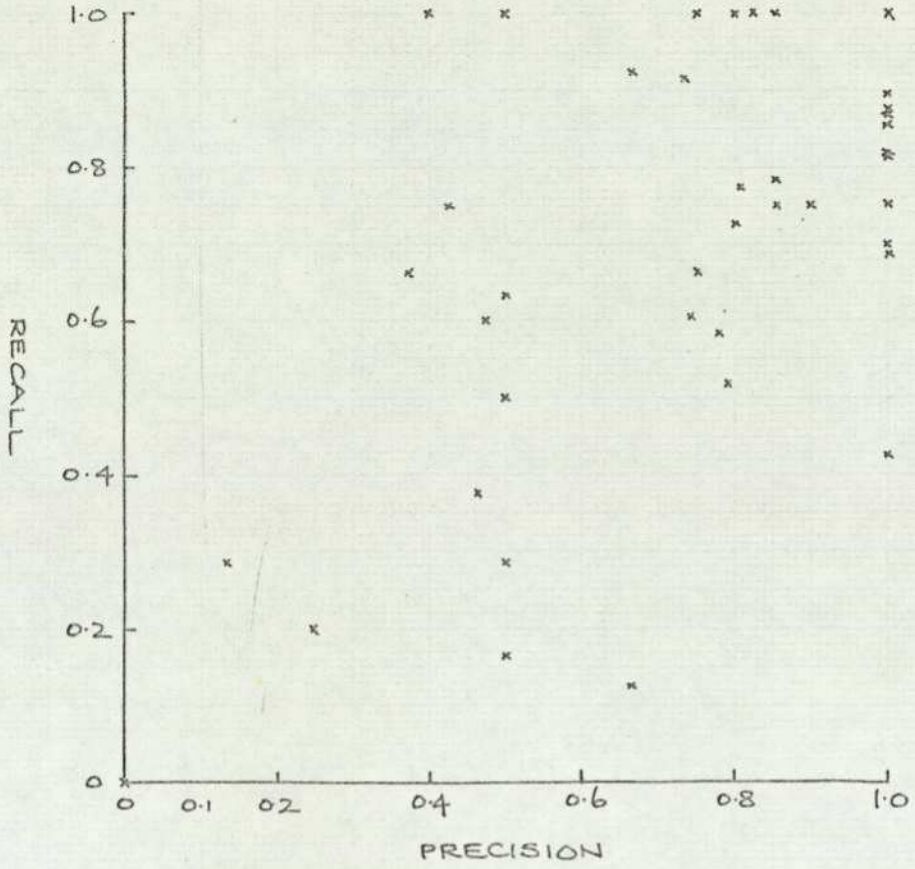


Recall - precision diagram (set 072)

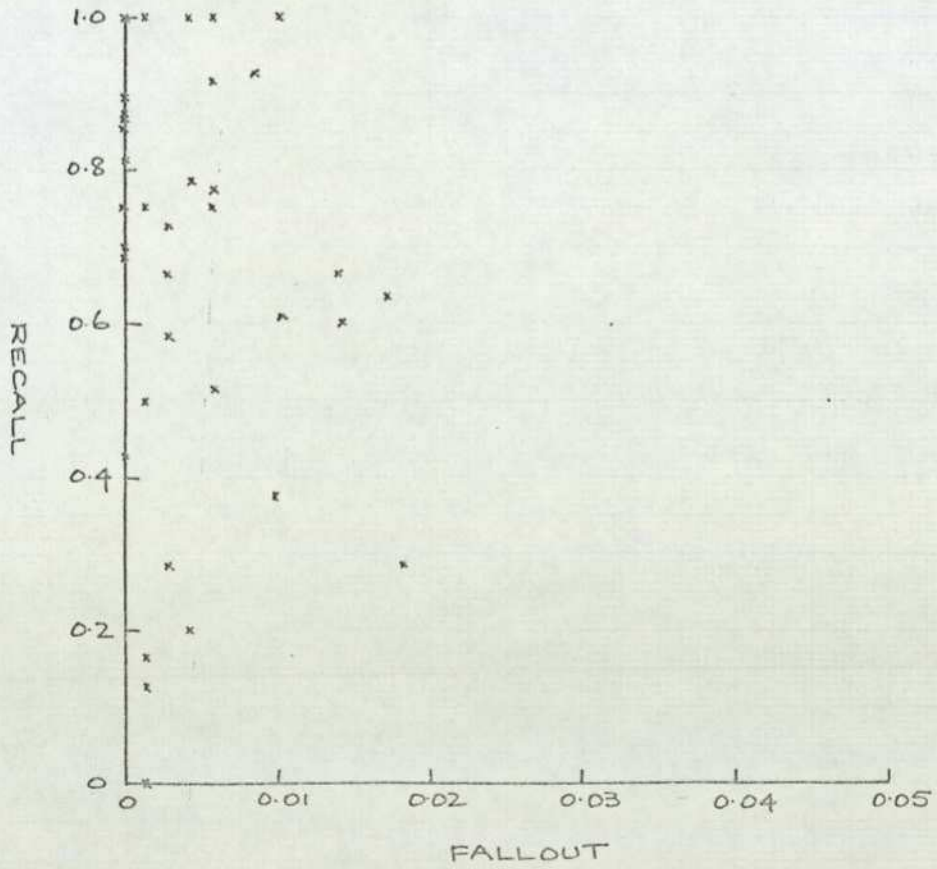


Recall - fallout diagram (set 072)

FIGURE 8

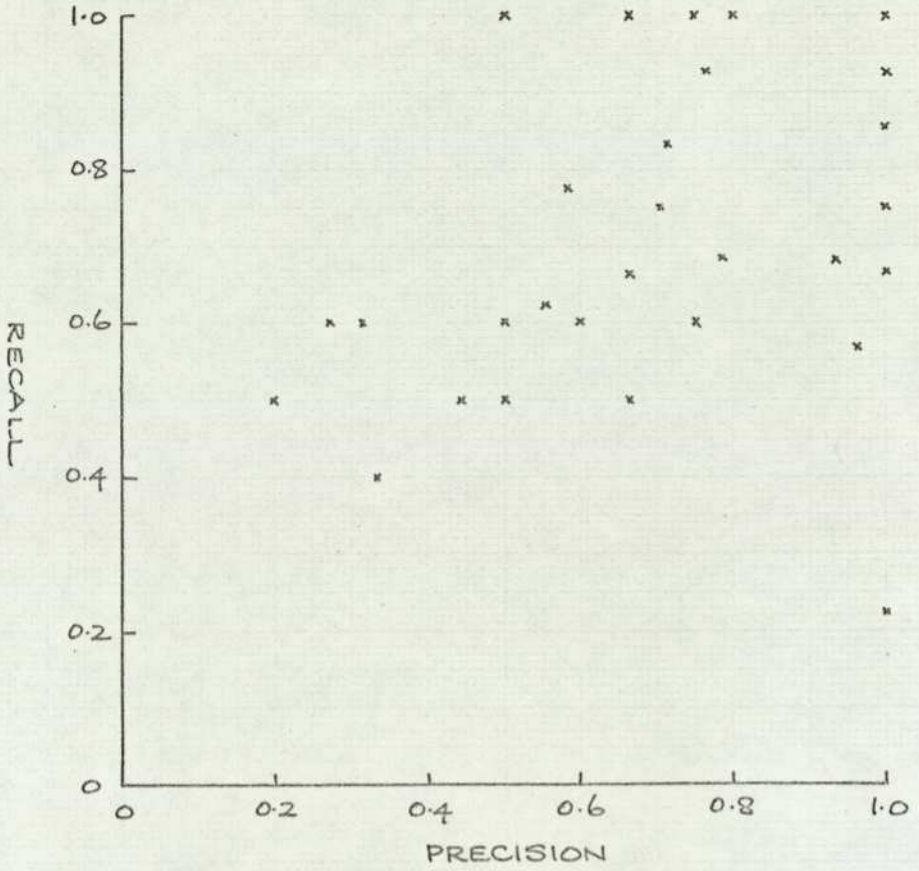


Recall – precision diagram (set 073)

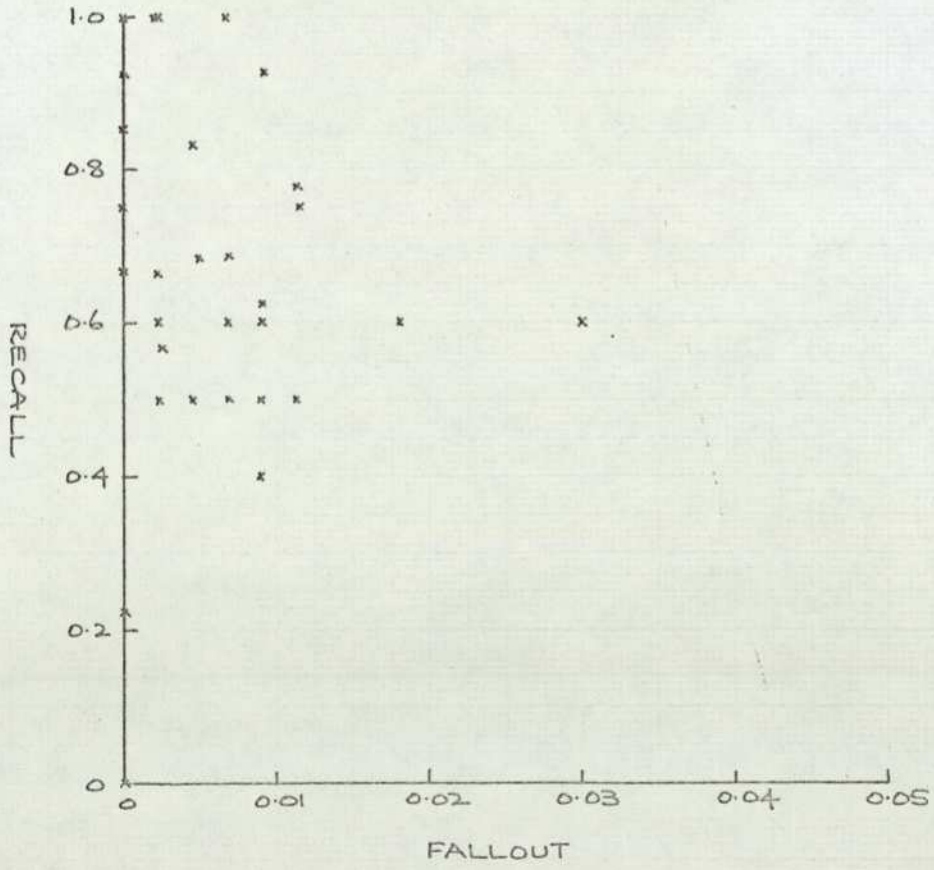


Recall – fallout diagram (set 073)

FIGURE 9

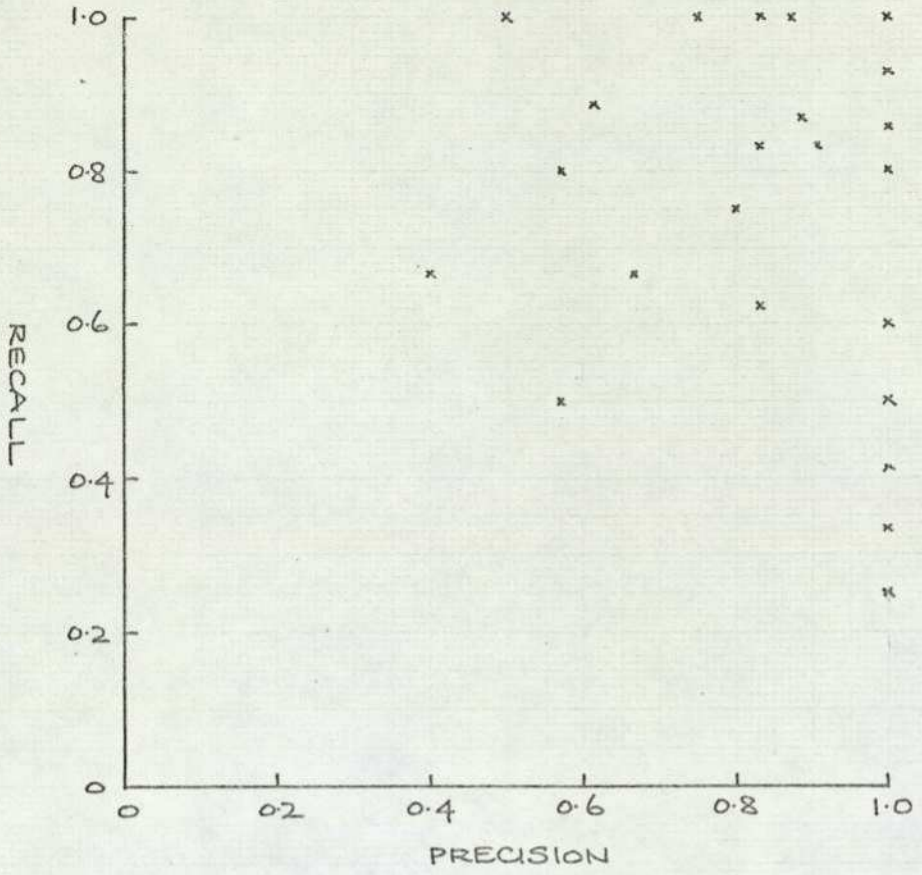


Recall - precision diagram (set 083)

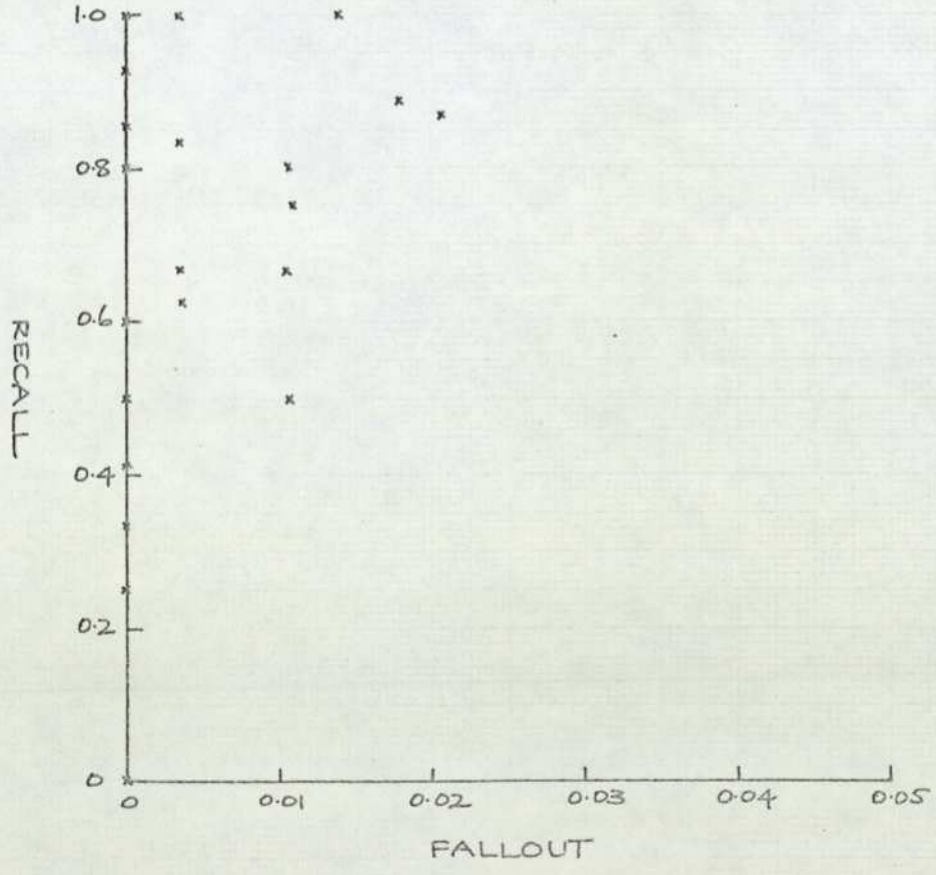


Recall - fallout diagram (set 083)

FIGURE 10

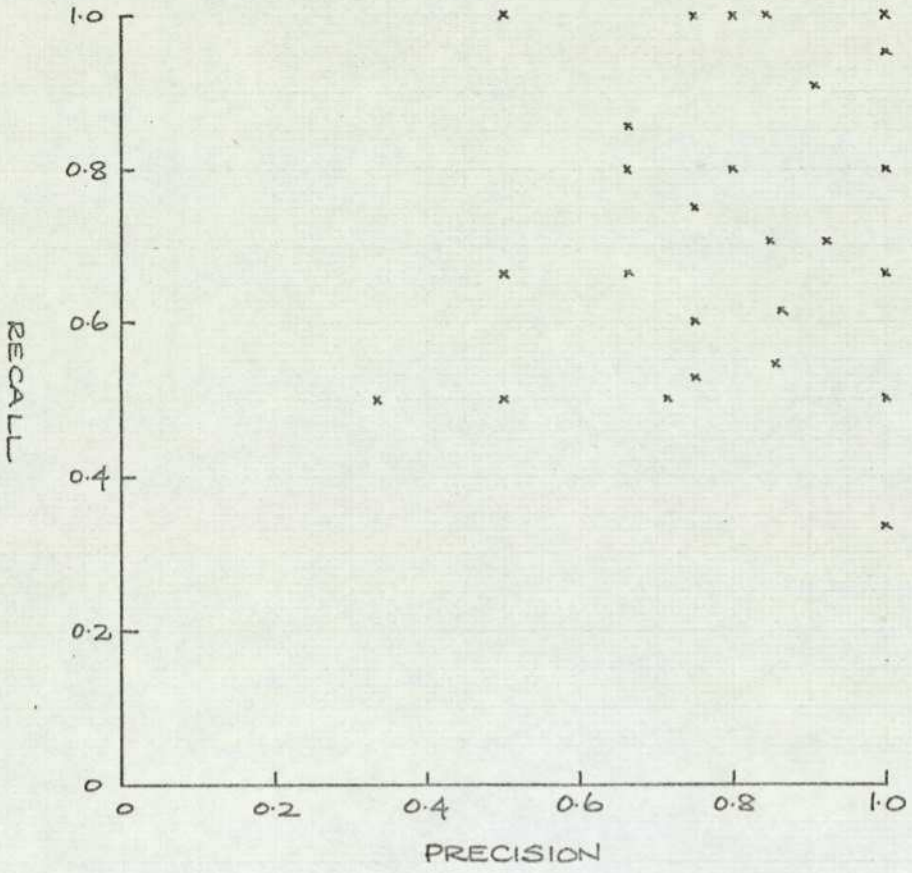


Recall – precision diagram (set 093)

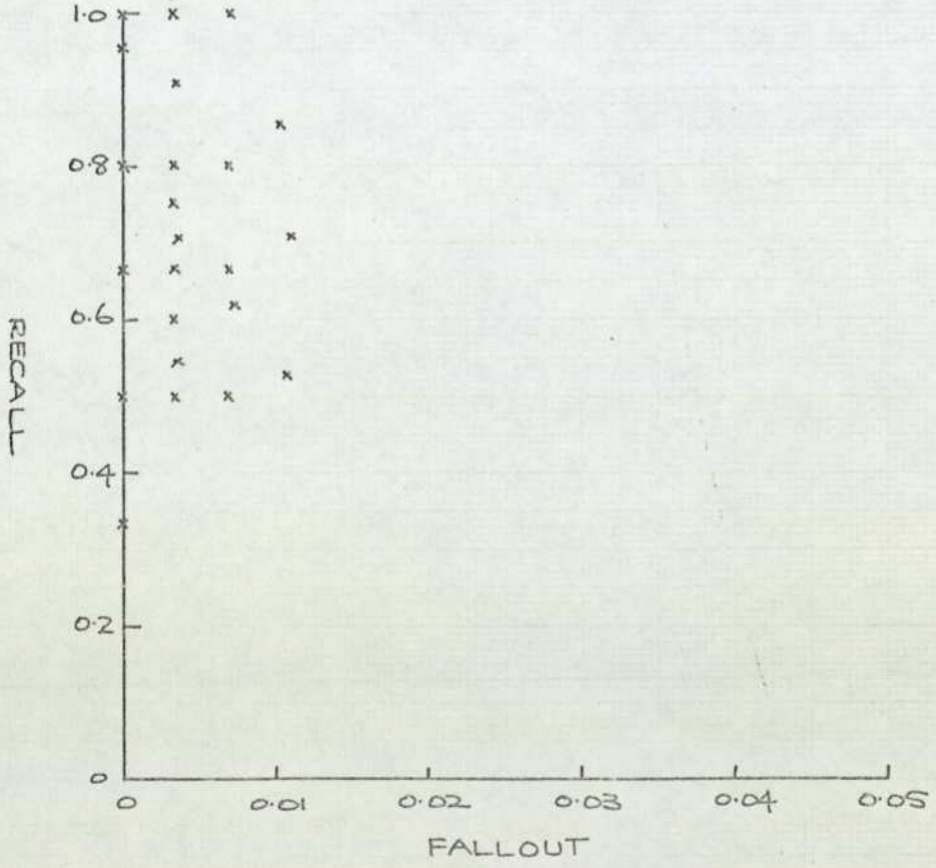


Recall - fallout diagram (set 093)

FIGURE 11

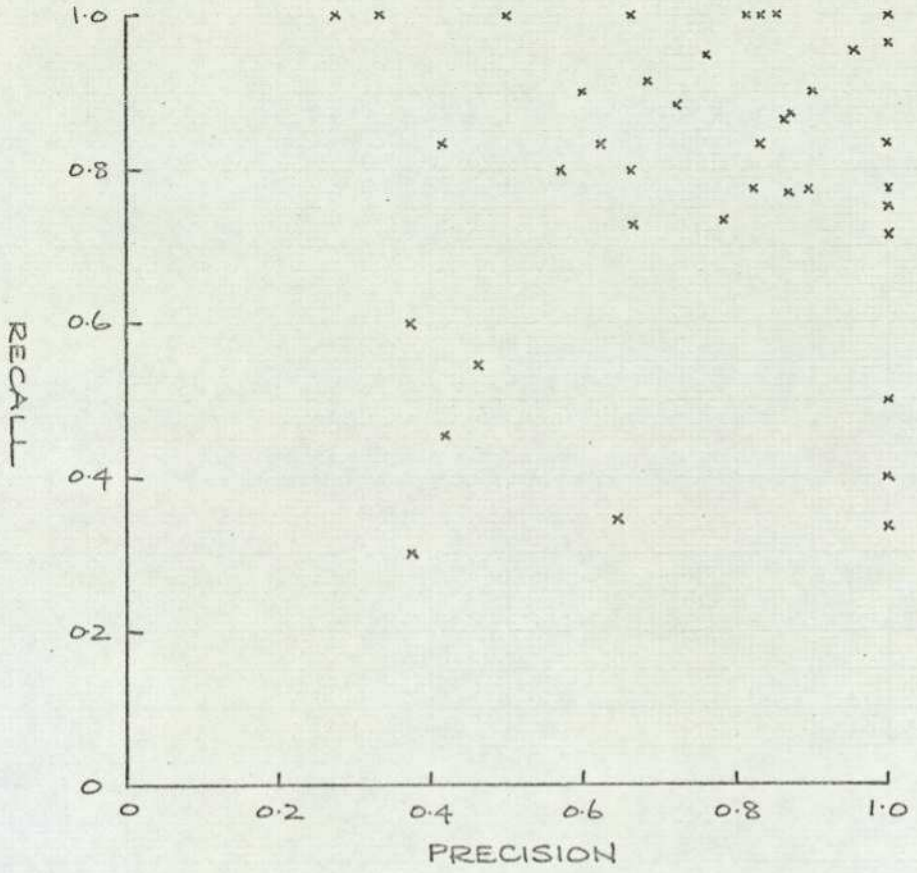


Recall - precision diagram (set 103)

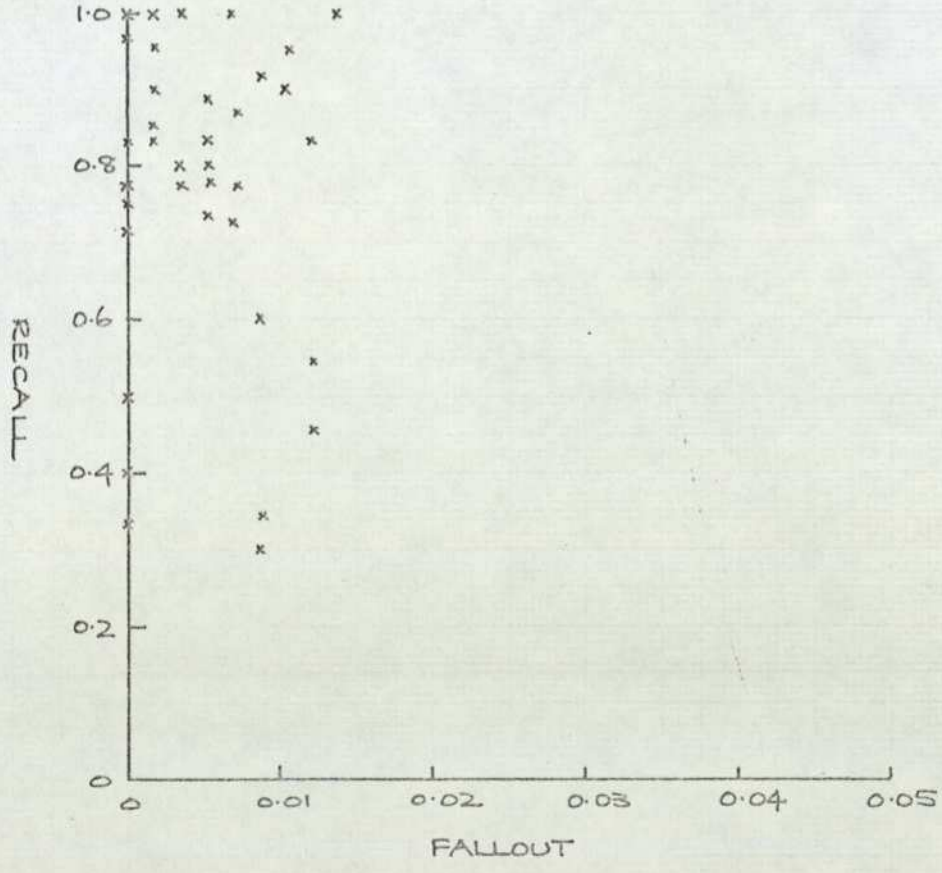


Recall - fallout diagram (set 103)

FIGURE 12

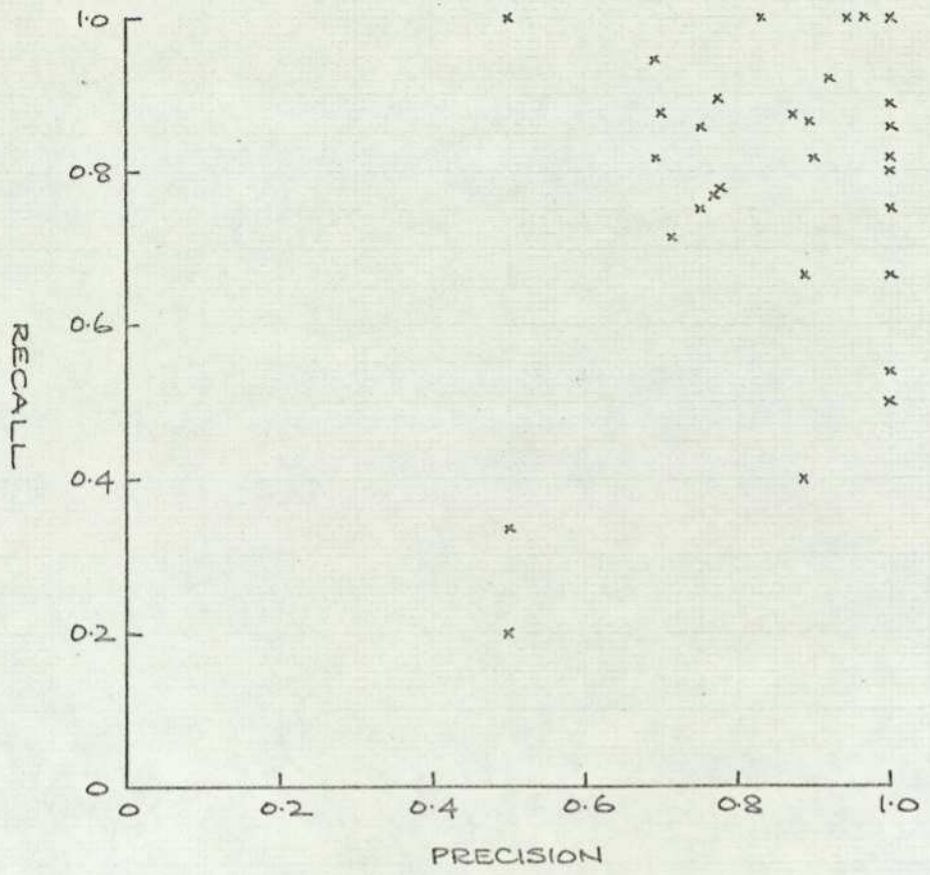


Recall - precision diagram (set 113)

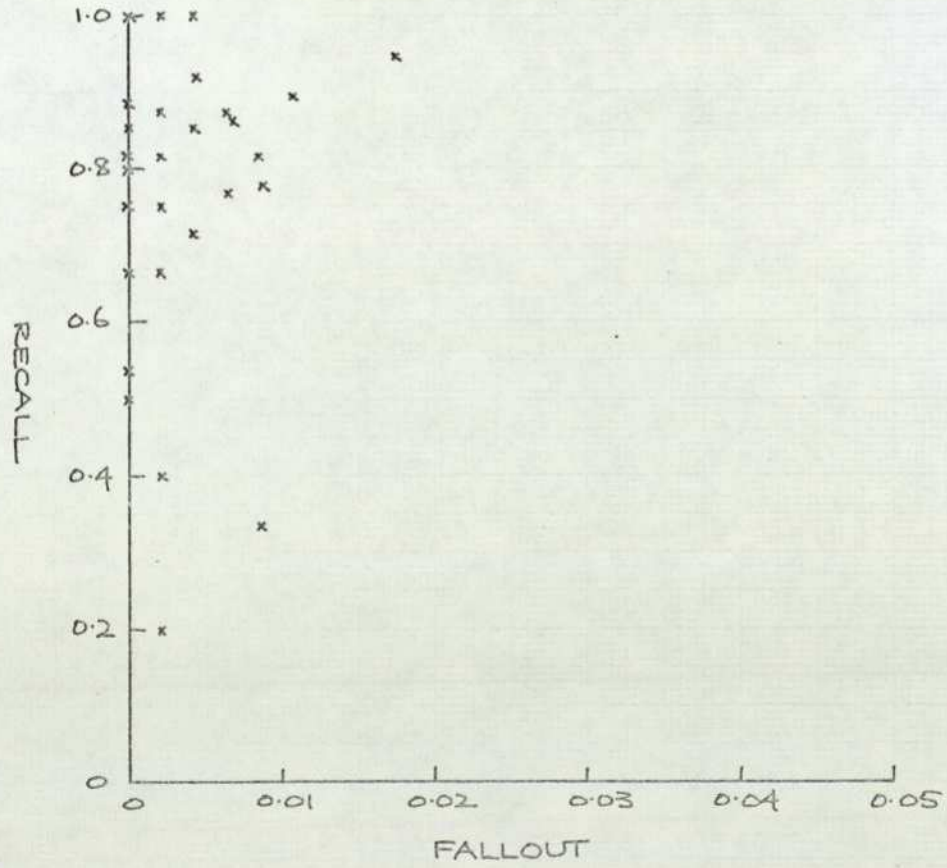


Recall - fallout diagram (set 113)

FIGURE 13



Recall – precision diagram (set 123)



Recall - fallout diagram (set 123)

FIGURE 14

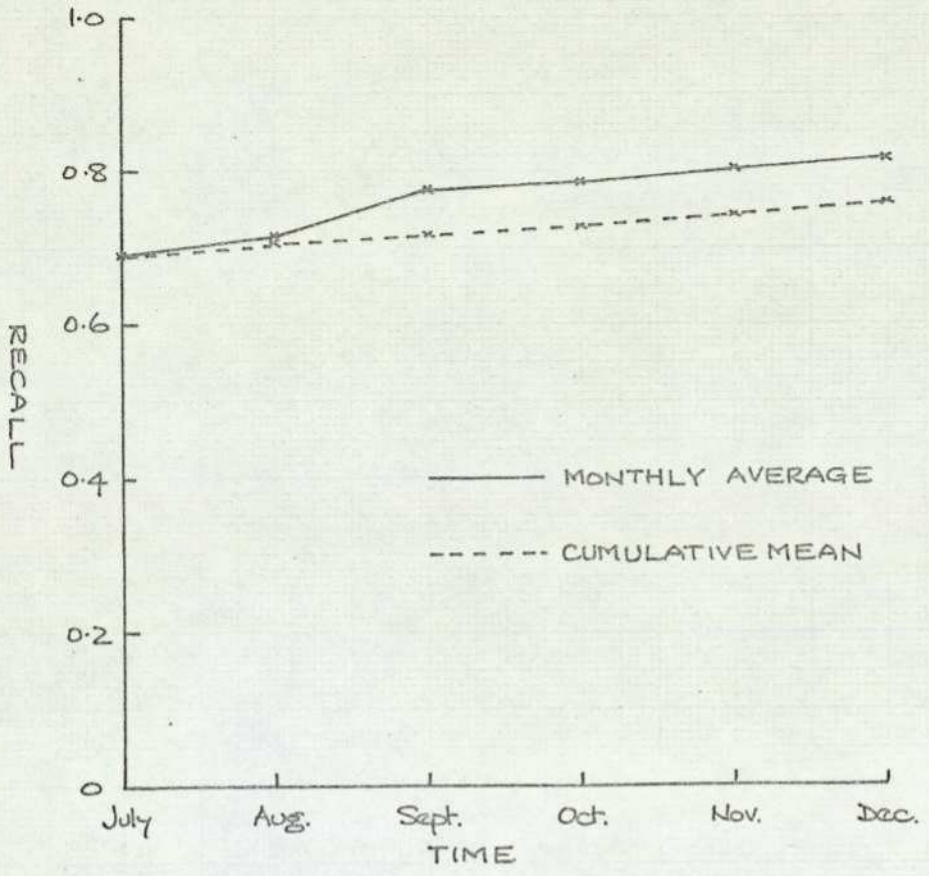


FIGURE 15 Monthly average and cumulative mean of recall (set XY3)

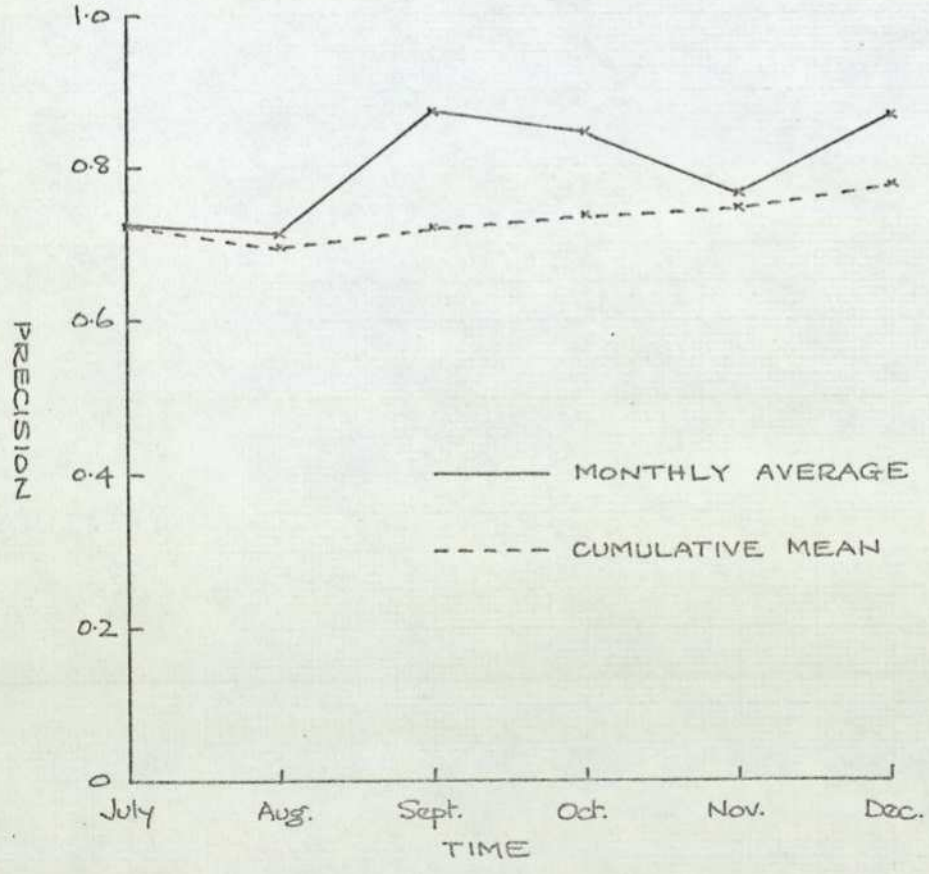


FIGURE 16 Monthly average and cumulative mean of precision (set XY3)

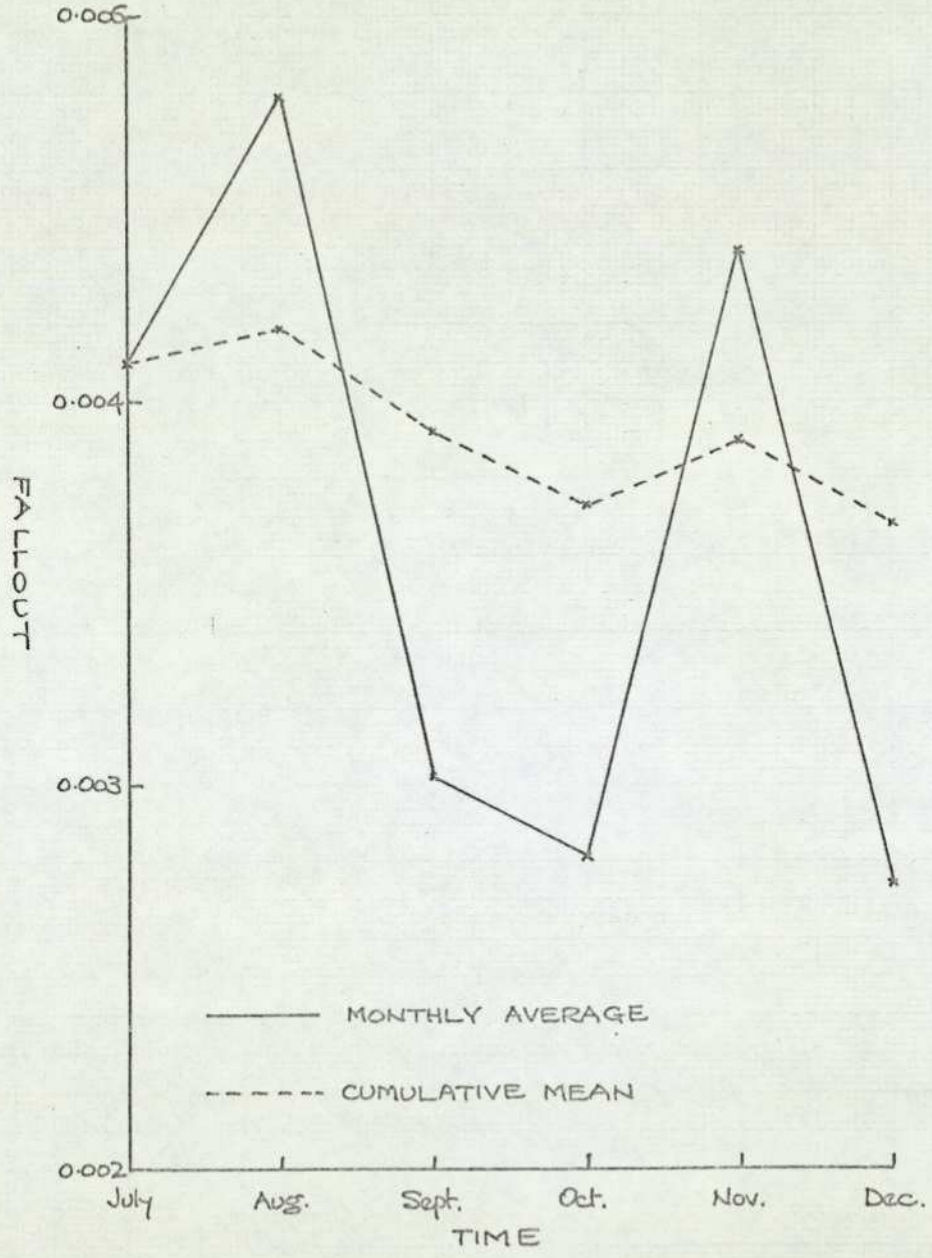
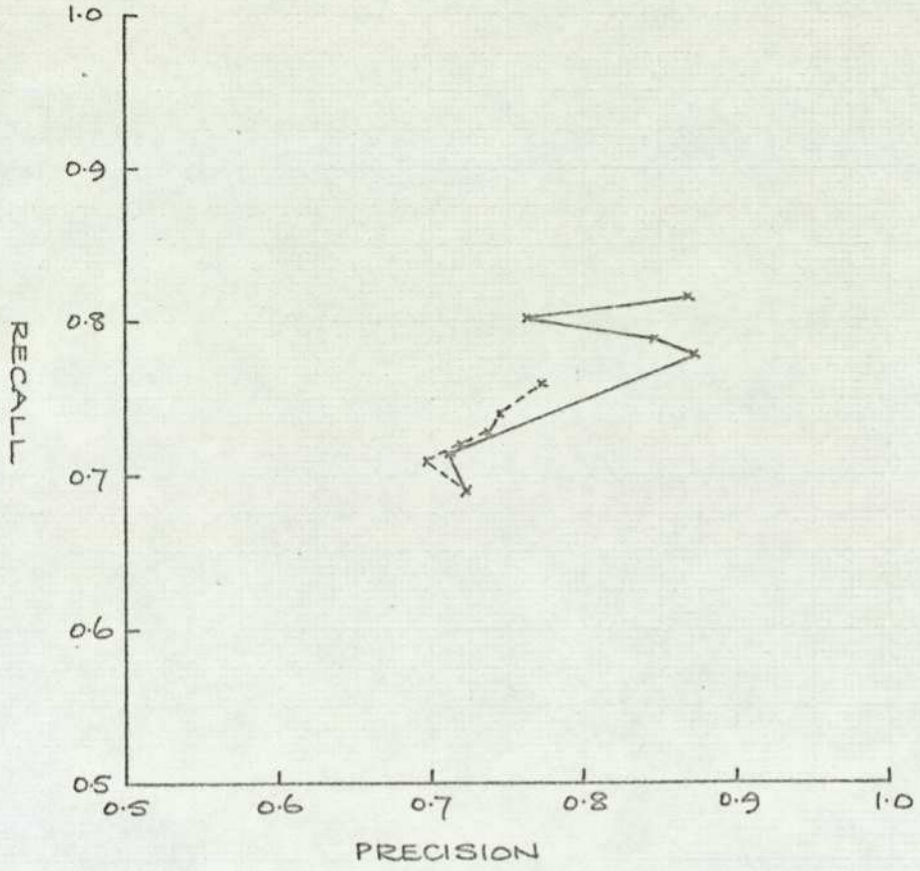
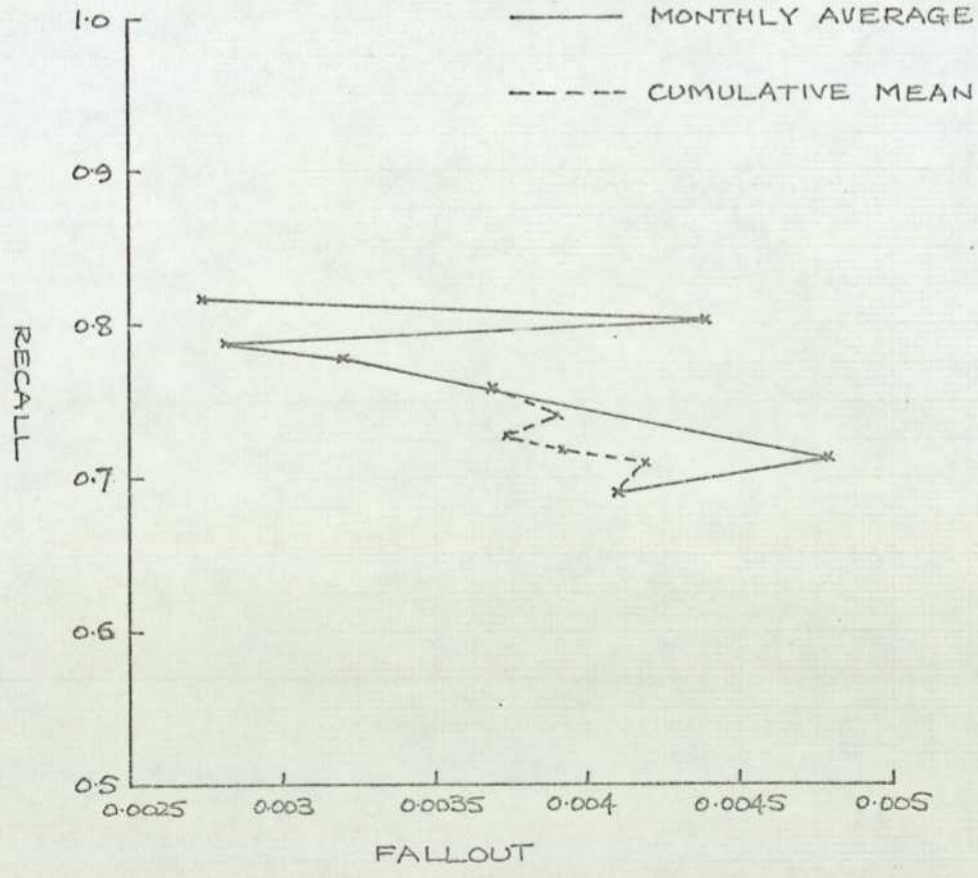


FIGURE 17 Monthly average and cumulative mean for fallout (set XY3)

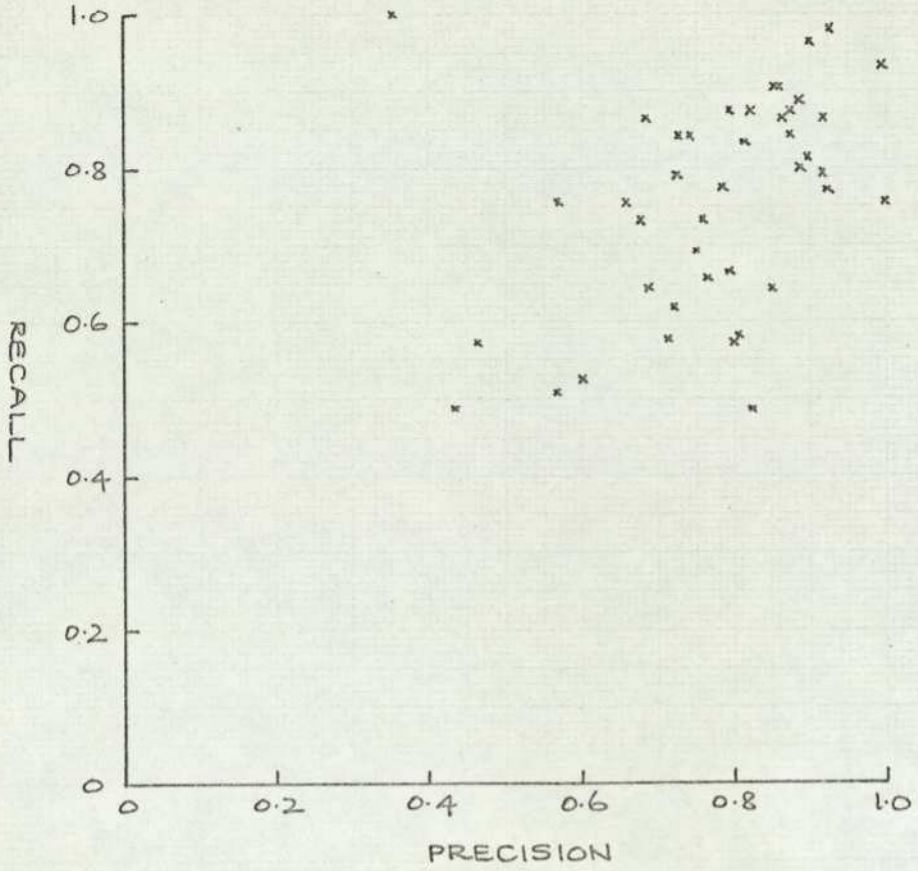


Recall – precision diagram (monthly average and cumulative mean)

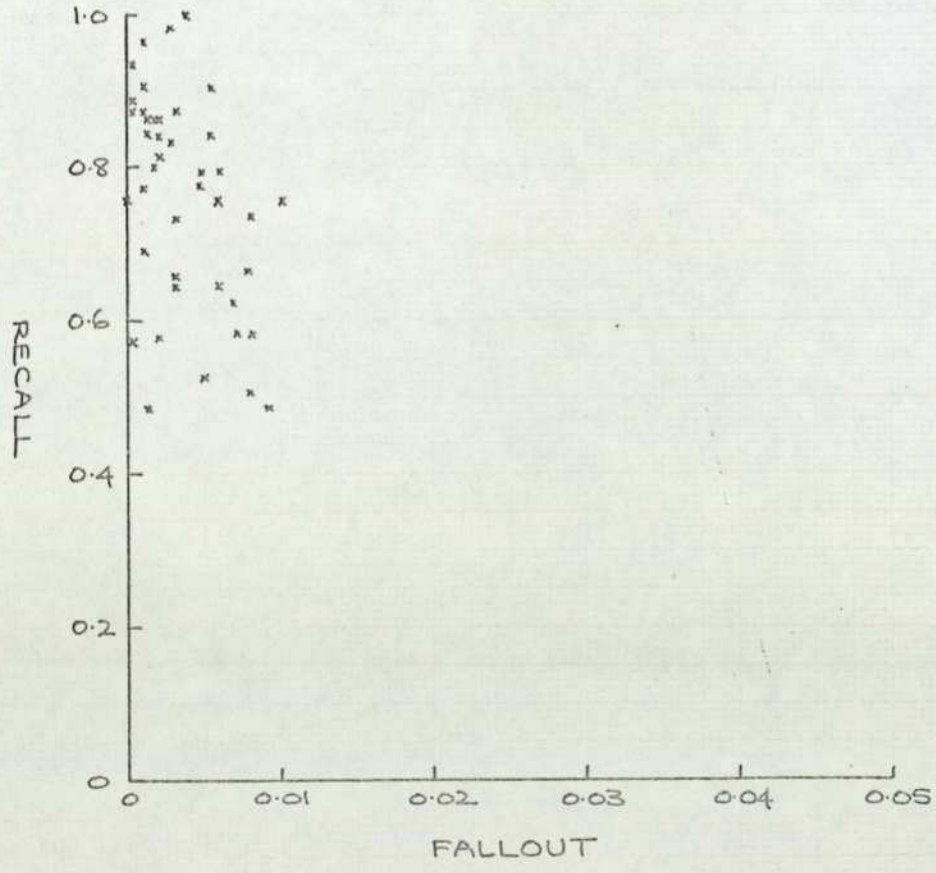


Recall - fallout diagram (monthly average and cumulative mean)

FIGURE 18



Recall-precision diagram (set 712)



Recall-fallout diagram (set 712)

FIGURE 19

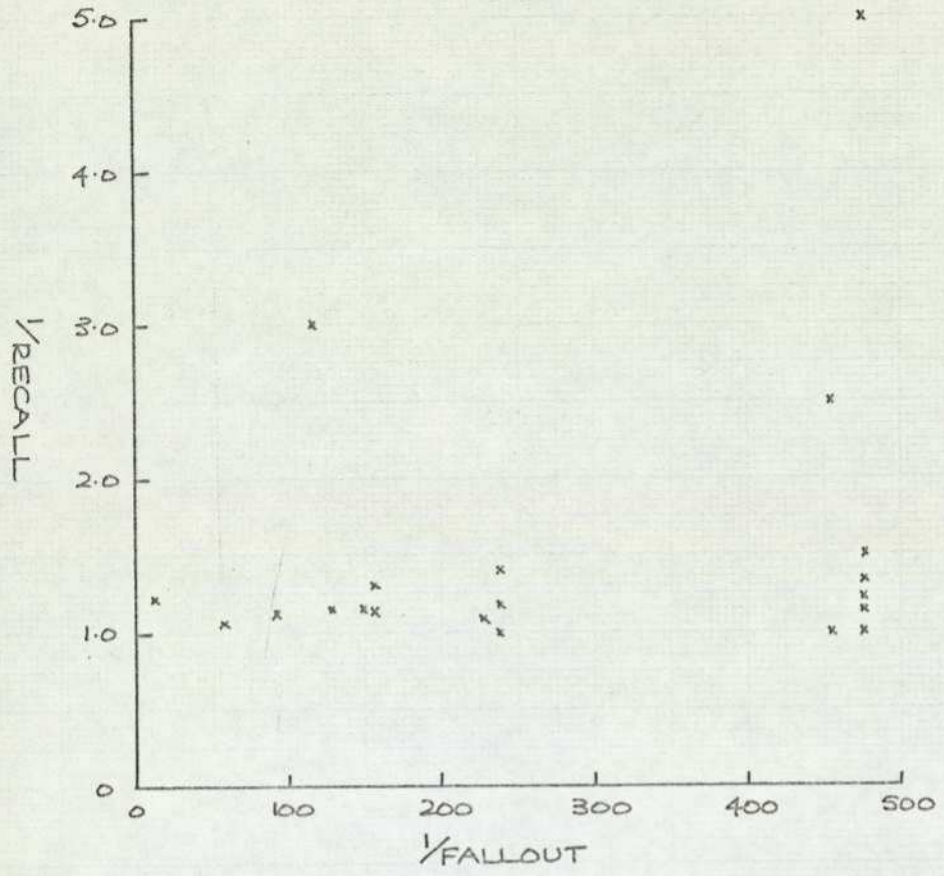


FIGURE 23 1/Recall - 1/Fallout graph (set 123)

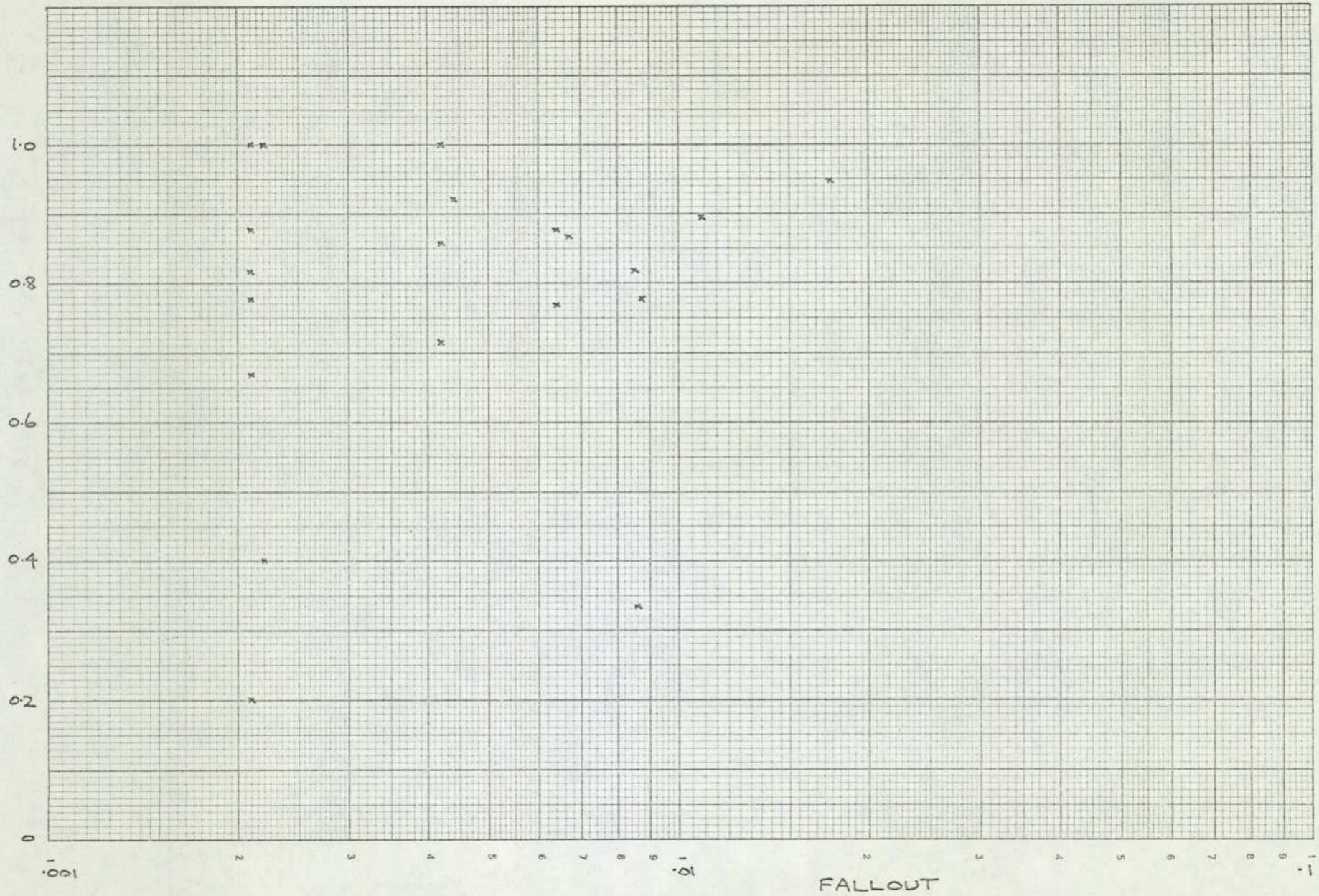
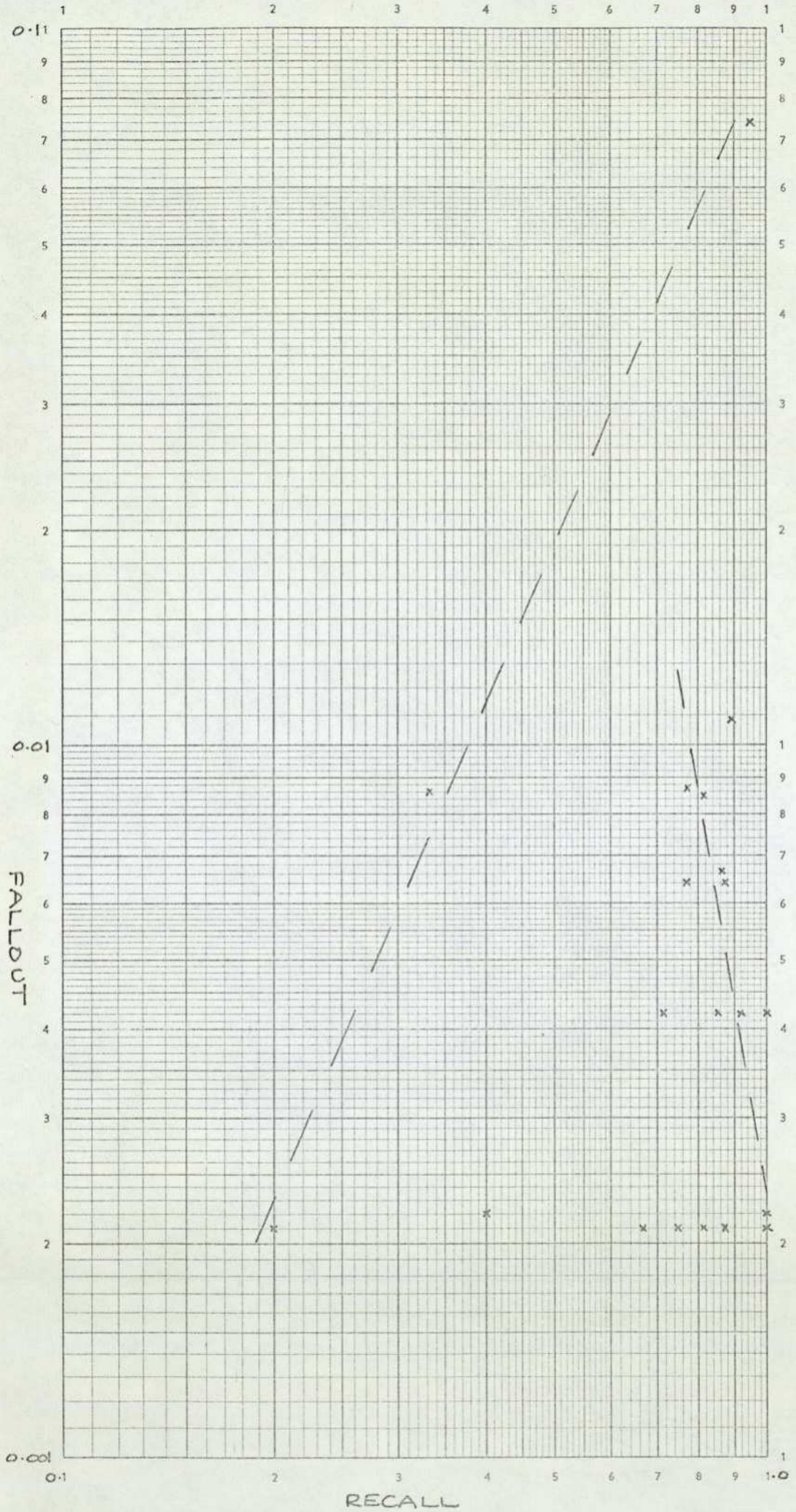


FIGURE 24 Recall – log Fallout graph (set 123)

FIGURE 25 Log Recall - log Fallout graph (set 123)



araph paper When ordering please quote Prefix CS, CB, TS or TB followed by Graph Reference 5027
8 2 Cycles Prefix CS. Cartridge paper printed Sepia Prefix CB. Cartridge paper printed Blue
1 Cycle Prefix TS. Tracing paper printed Sepia Prefix TB. Tracing paper printed Blue

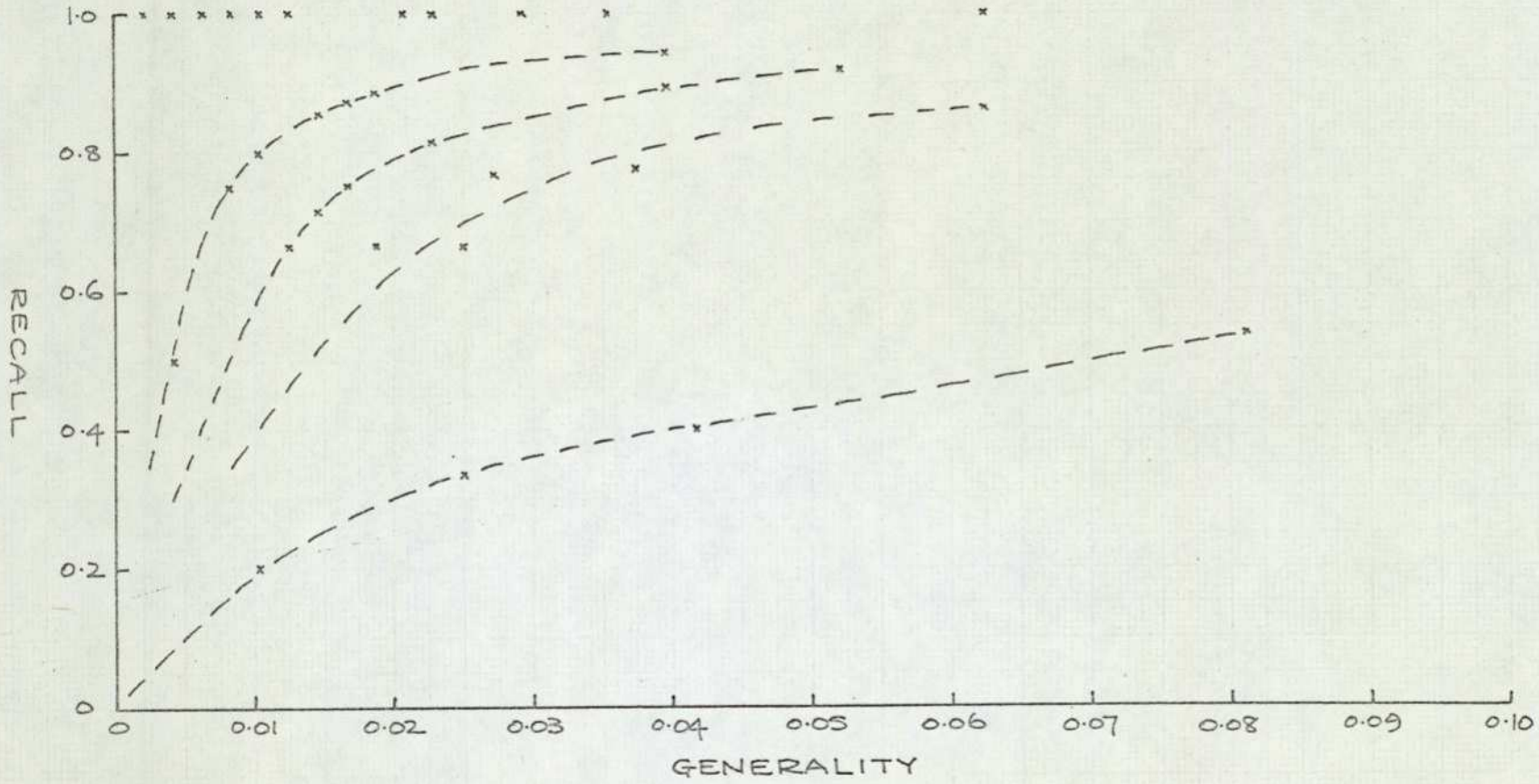
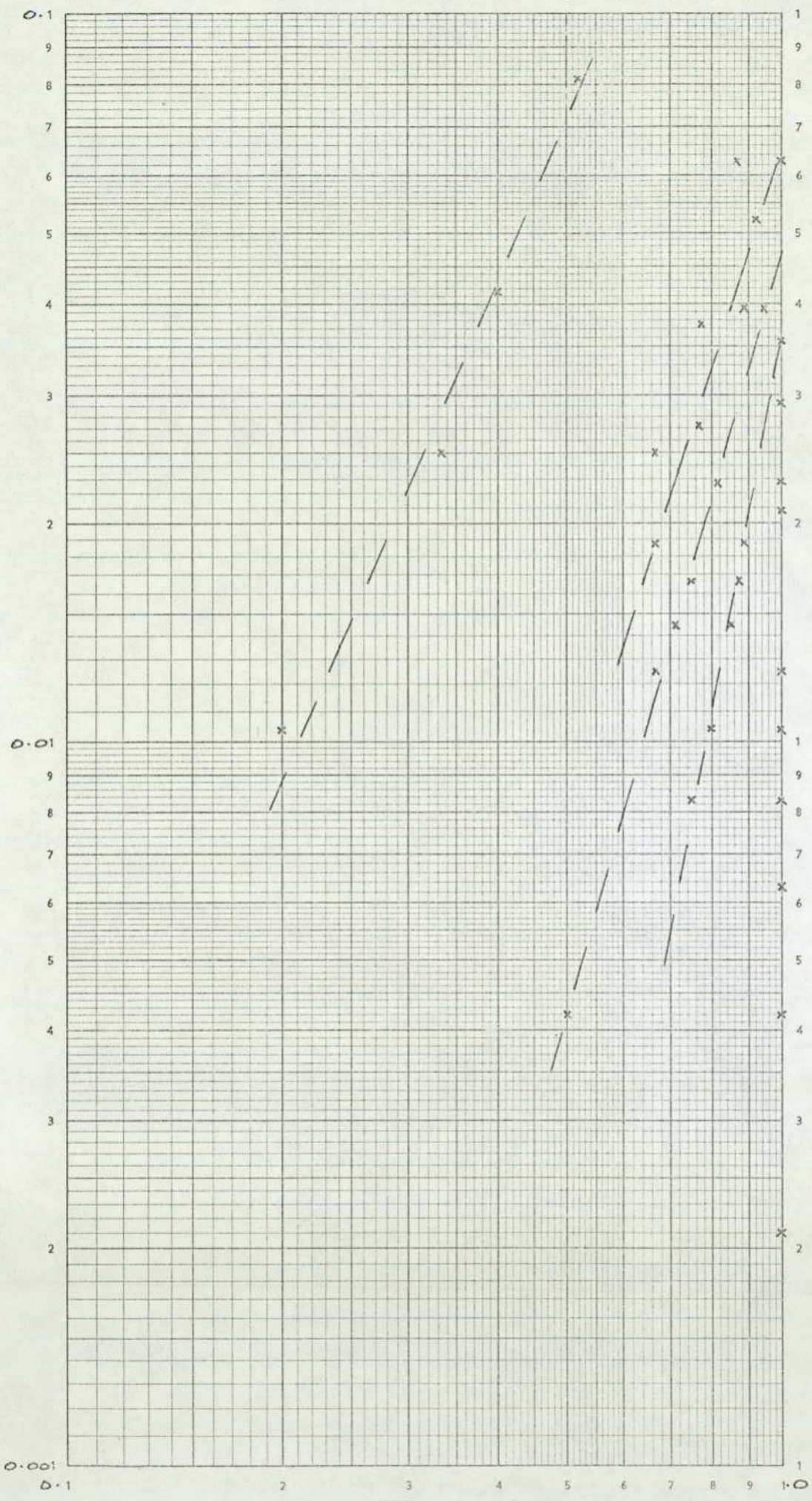


FIGURE 27 Recall - generality graph (set 123)

FIGURE 28 Log Recall – log Generality graph (set 123)



paph paper
g 2 Cycles
g 1 Cycle
When ordering please quote Prefix CS, CB, TS or TB followed by Graph Reference
Prefix CS. Cartridge paper printed Sepia
Prefix TS. Tracing paper printed Sepia
Prefix CB. Cartridge paper printed Blue
Prefix TB. Tracing paper printed Blue
5027

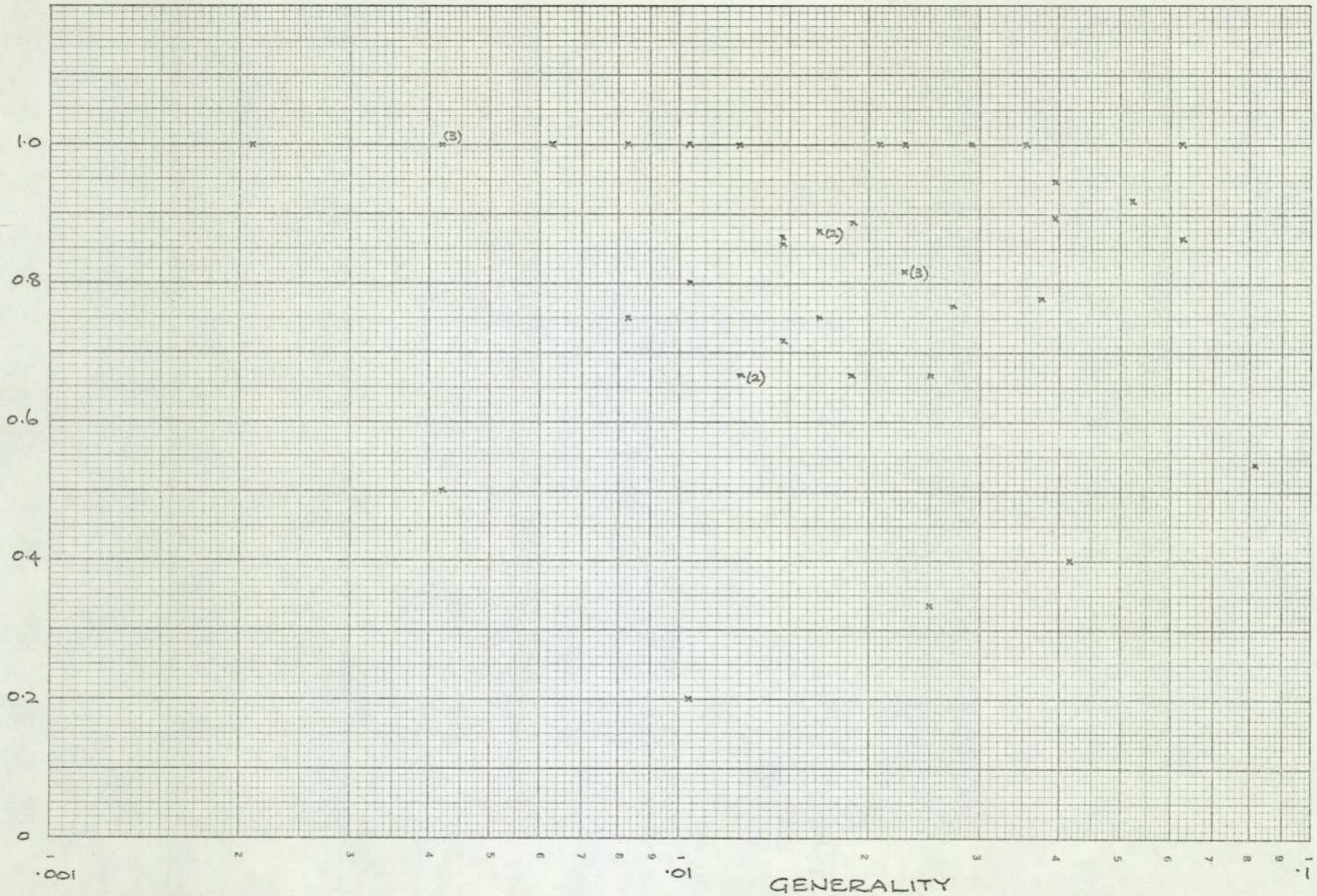


FIGURE 29 Recall – log Generality graph (set 123)

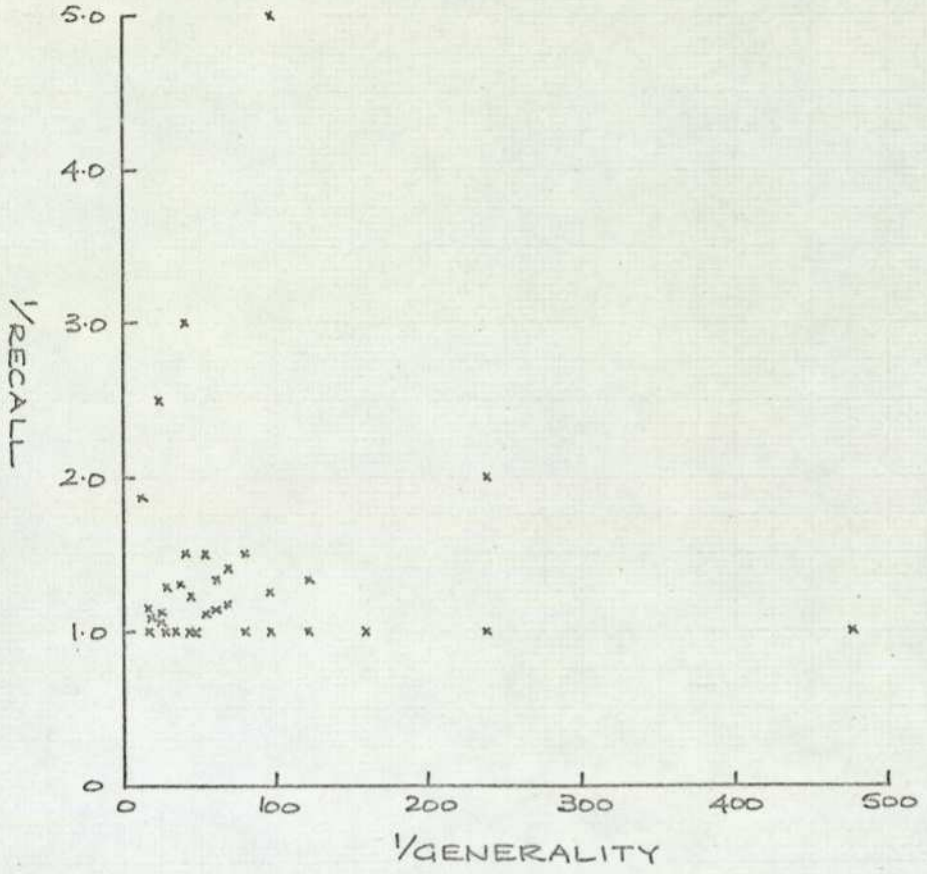


FIGURE 30 1/Recall – 1/Generality graph (set 123)

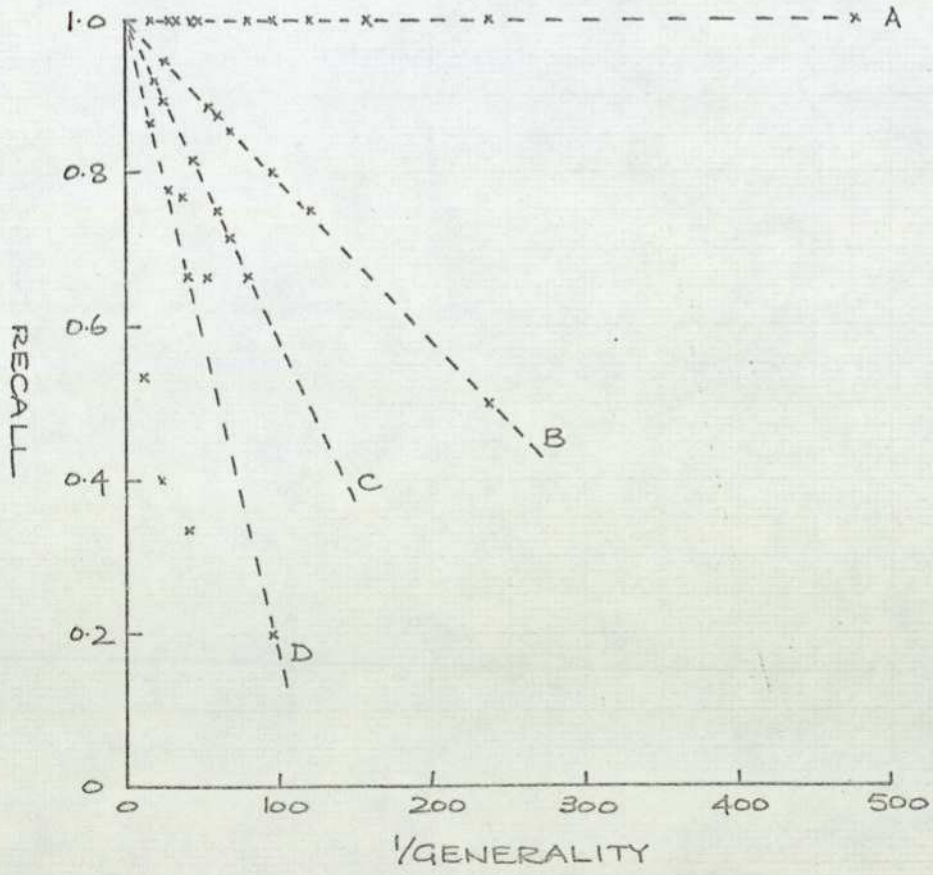


FIGURE 31 Recall – 1/Generality graph (set 123)

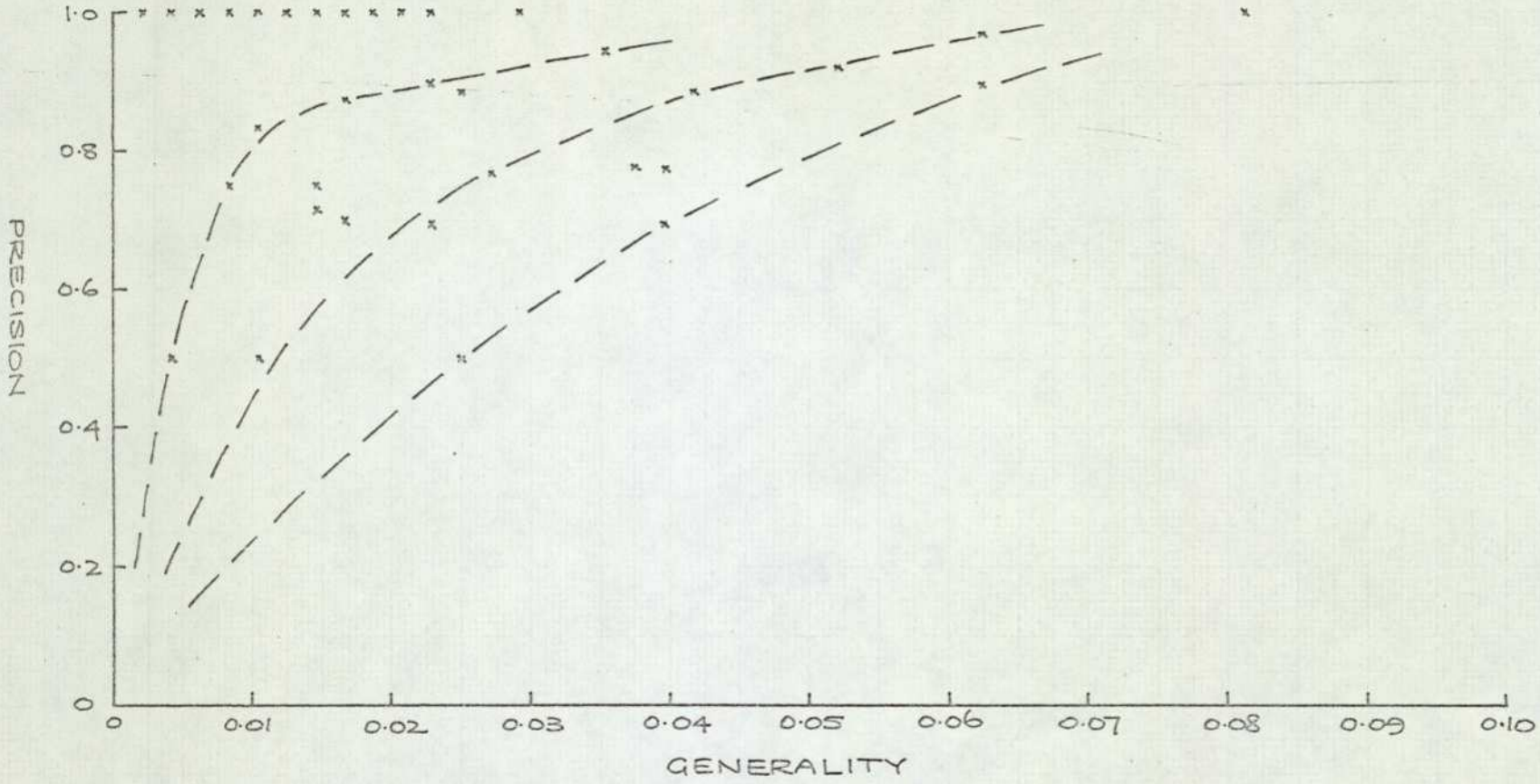
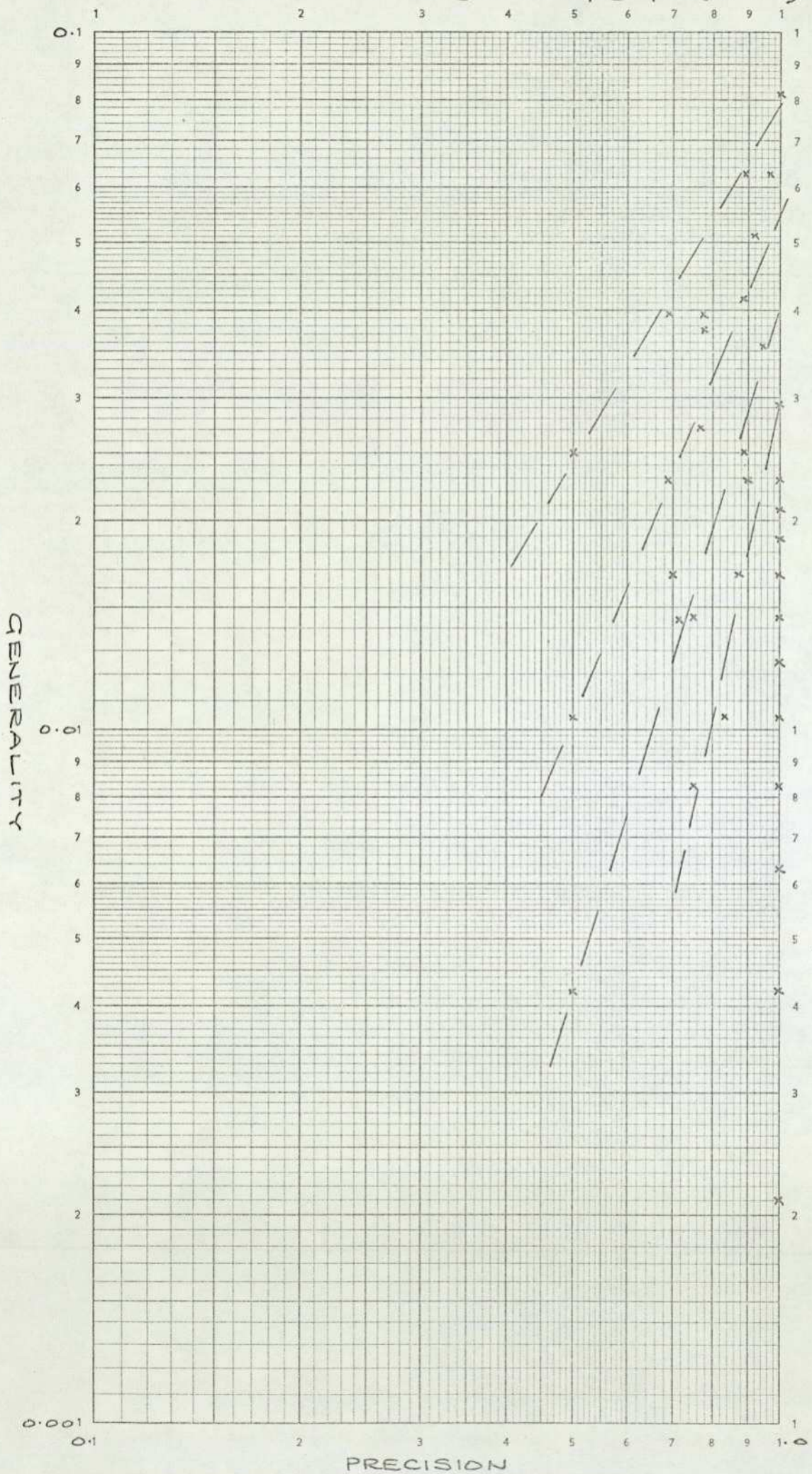


FIGURE 32 Precision – generality graph (set 123)

FIGURE 33 Log Precision - log Generality graph (set 123)



paph paper
2 2 Cycles
1 Cycle
When ordering please quote Prefix CS, CB, TS or TB followed by Graph Reference
Prefix CS. Cartridge paper printed Sepia
Prefix CB. Cartridge paper printed Blue
5027

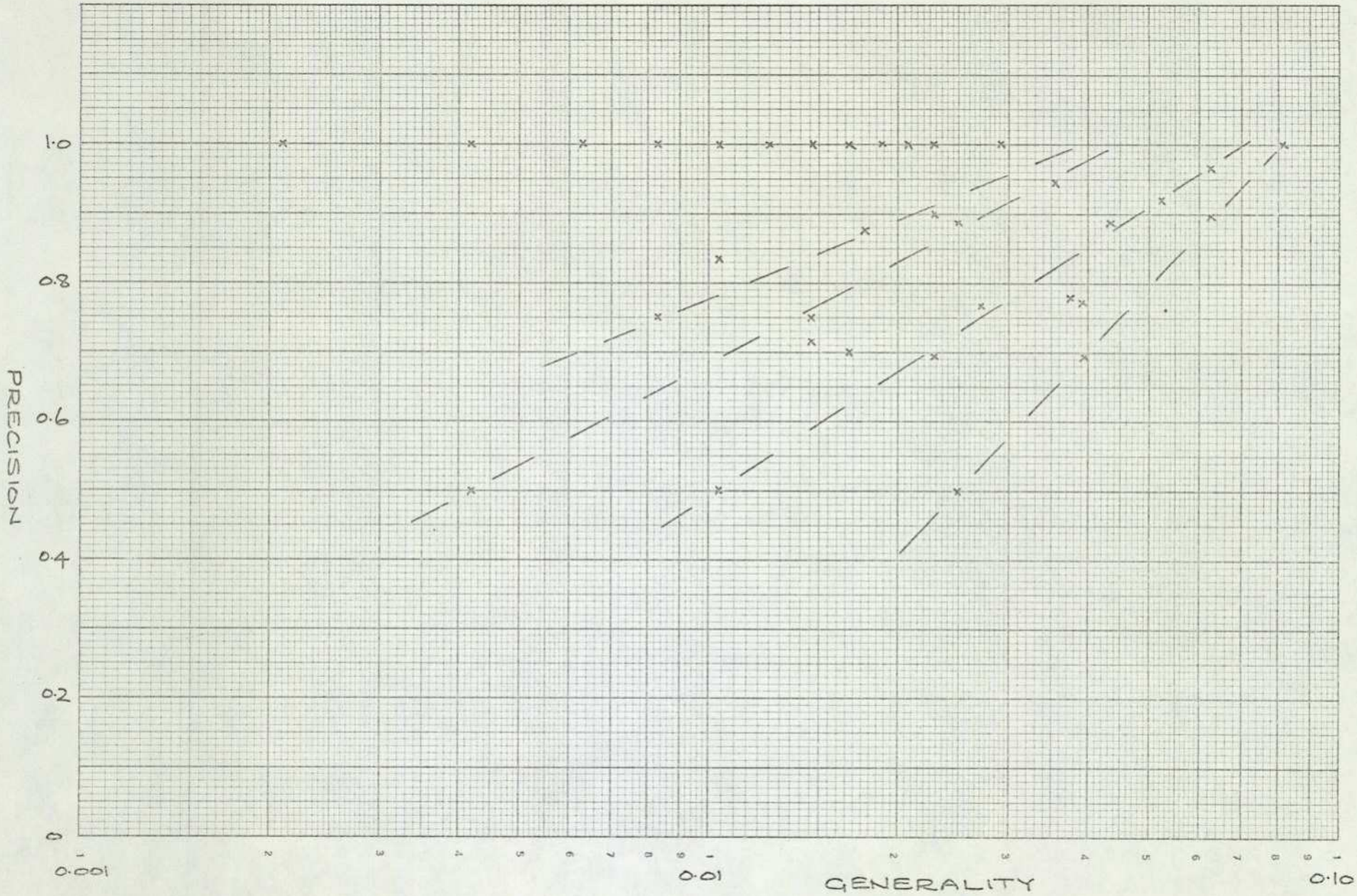


FIGURE 34 Precision – log Generality graph (set 123)

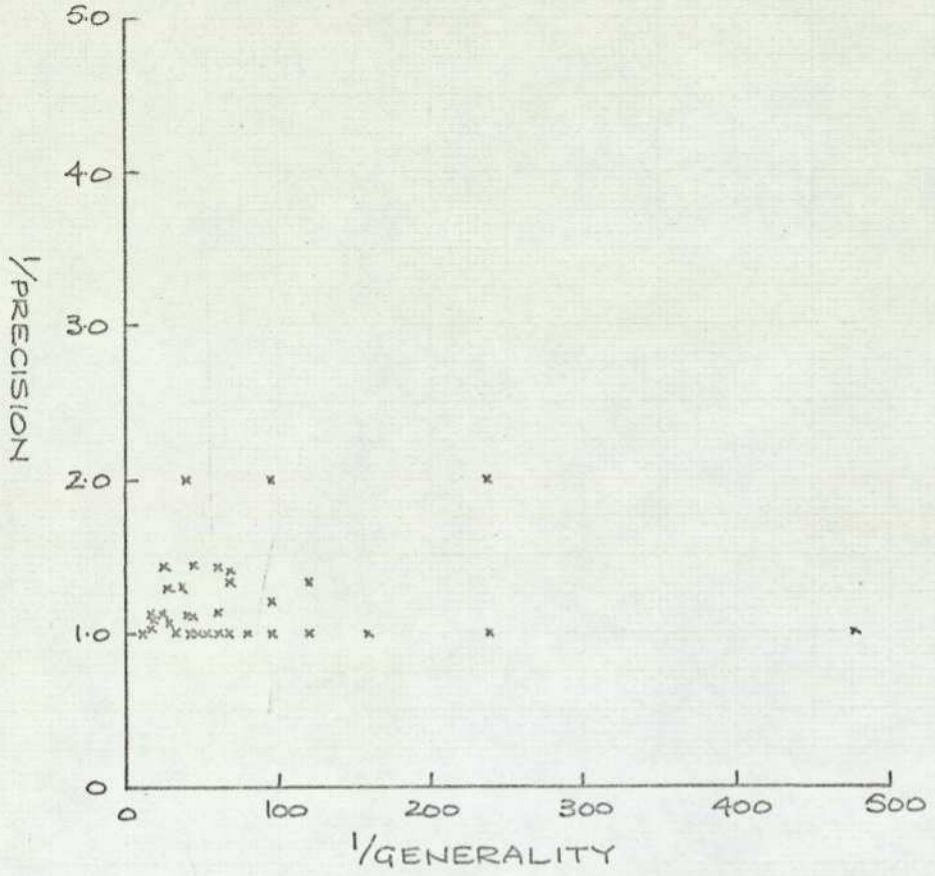


FIGURE 35 1/Precision – 1/Generality graph (set 123)

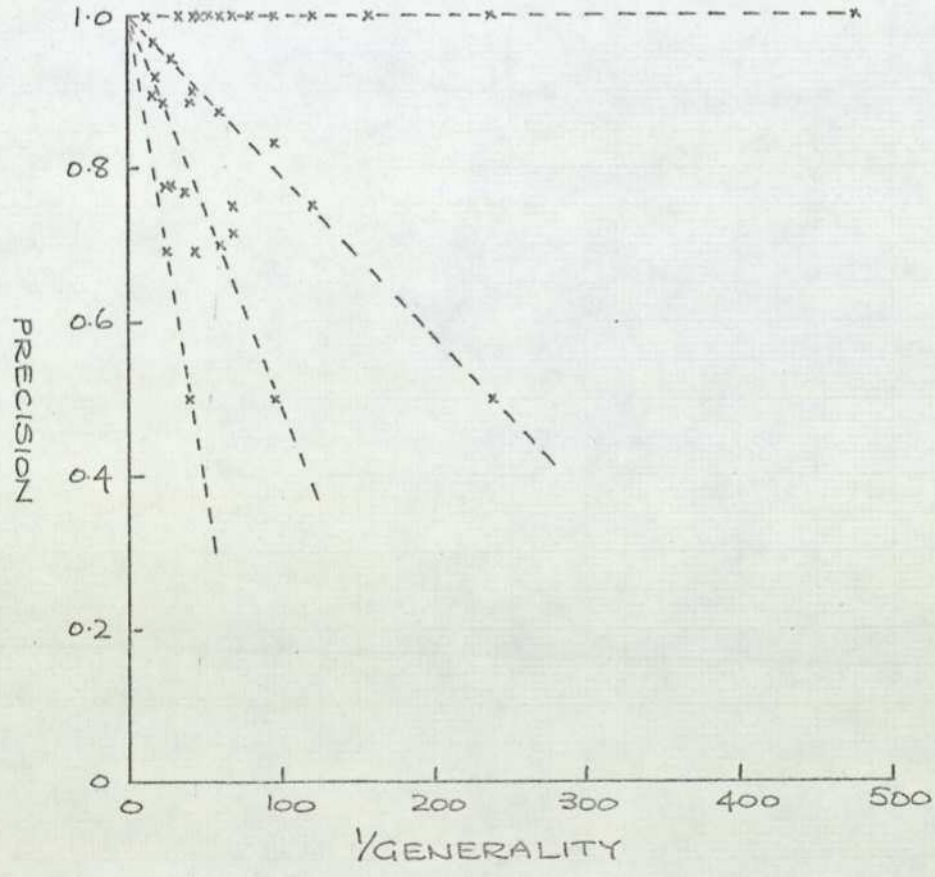


FIGURE 36 Precision – 1/Generality graph (set 123)

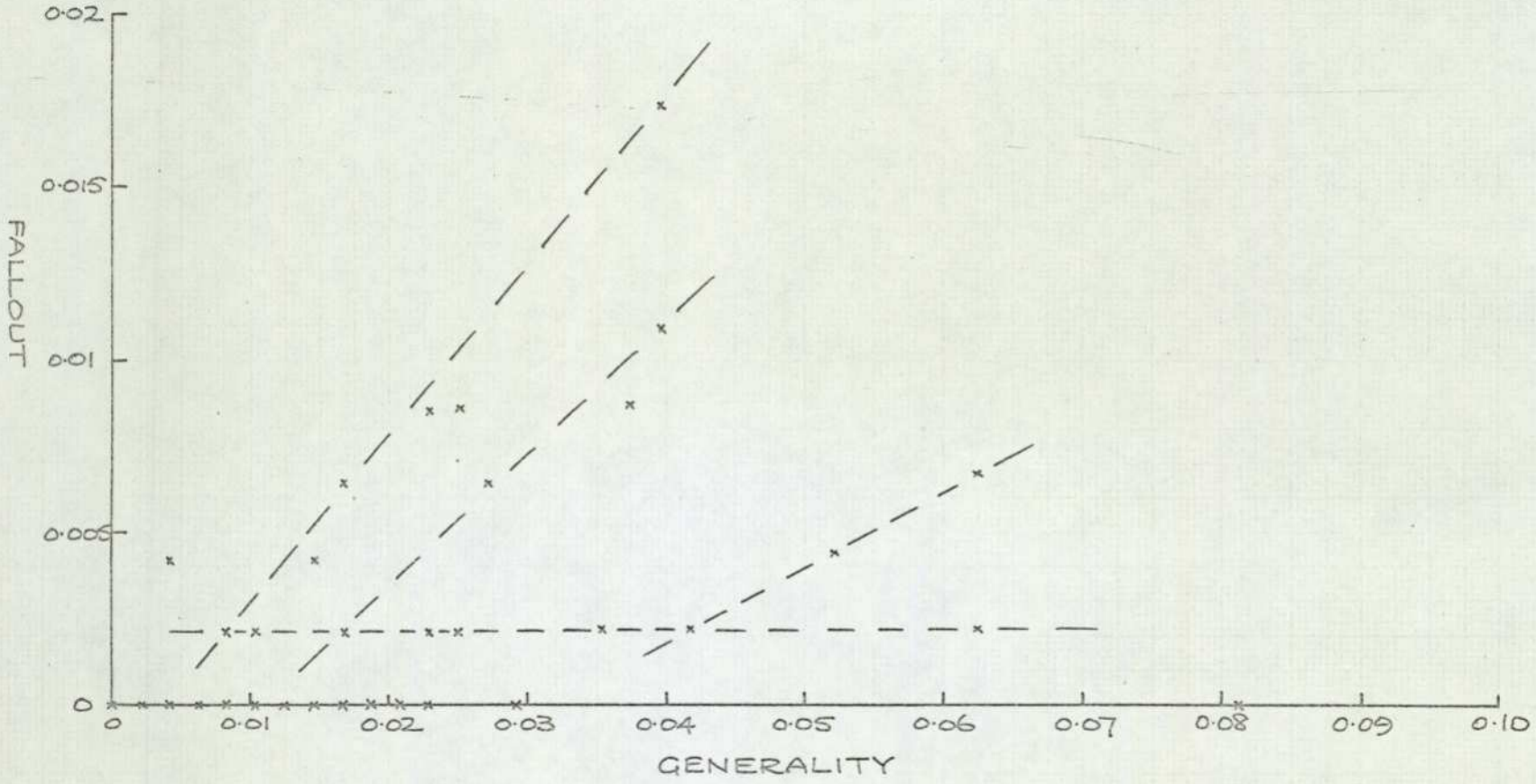
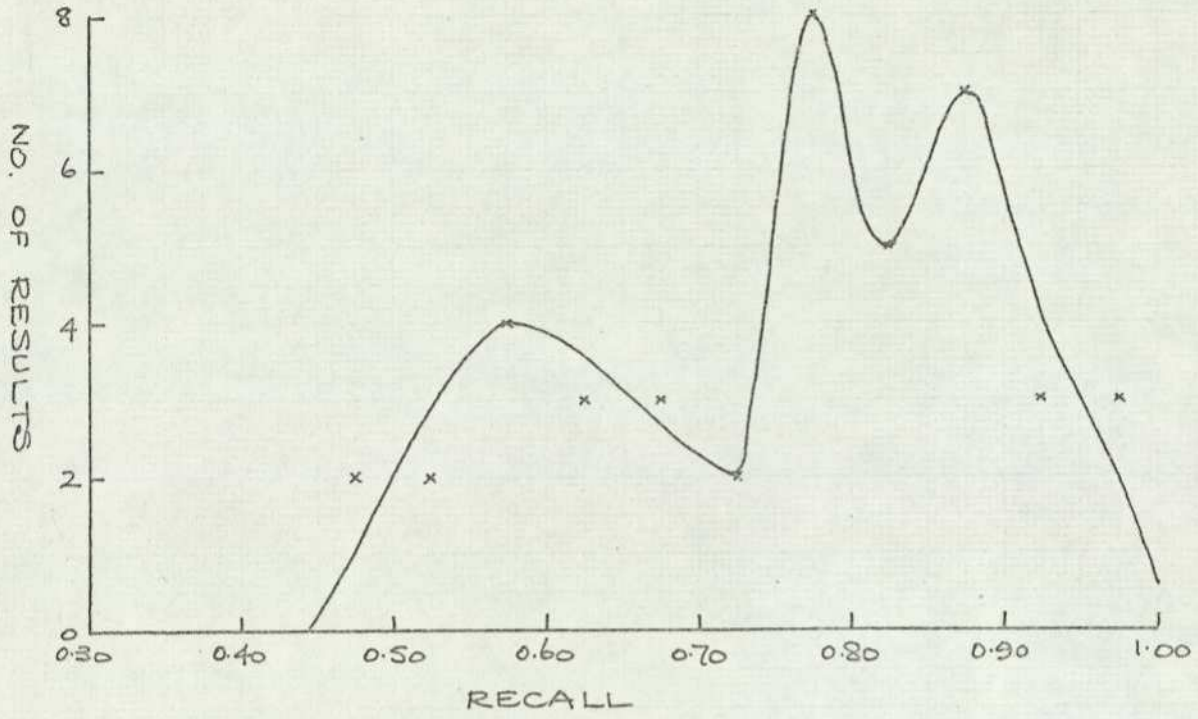
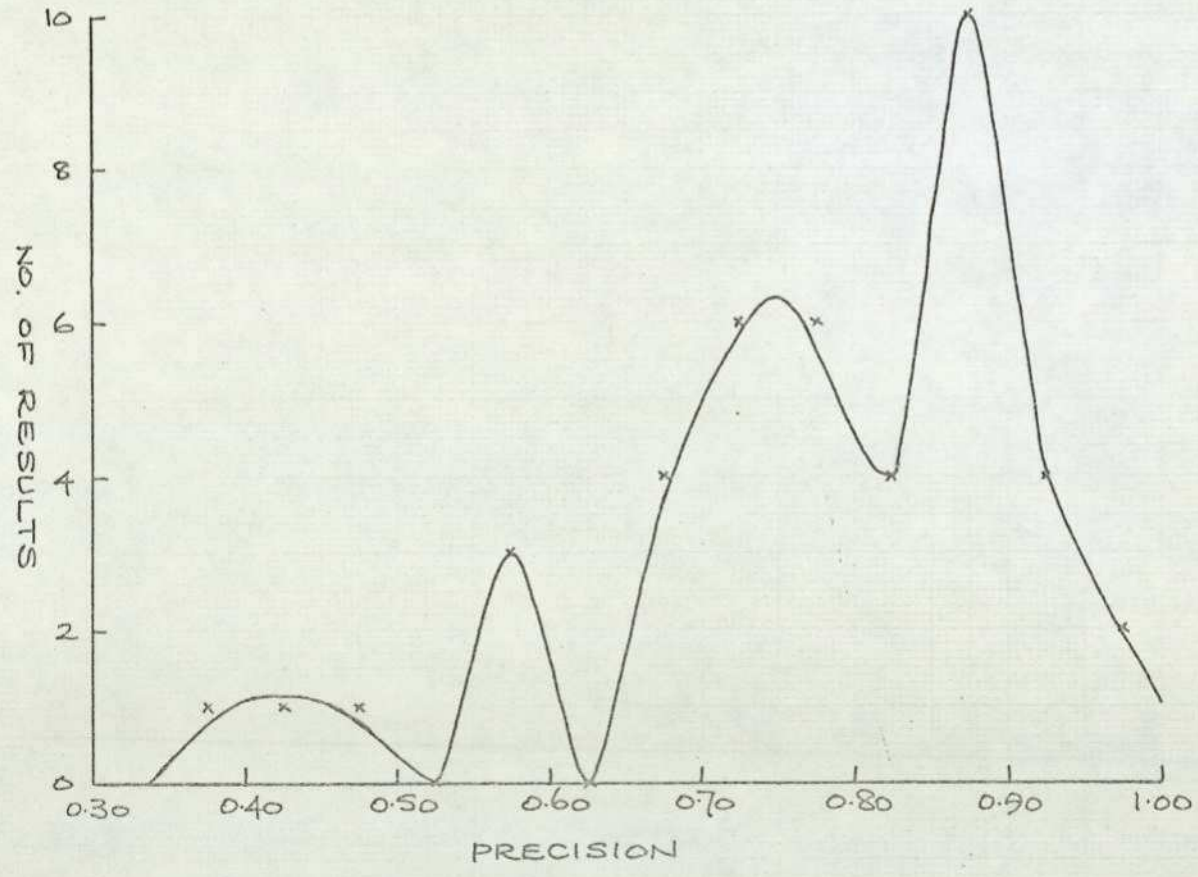


FIGURE 37 Fallout – generality graph (set 123)

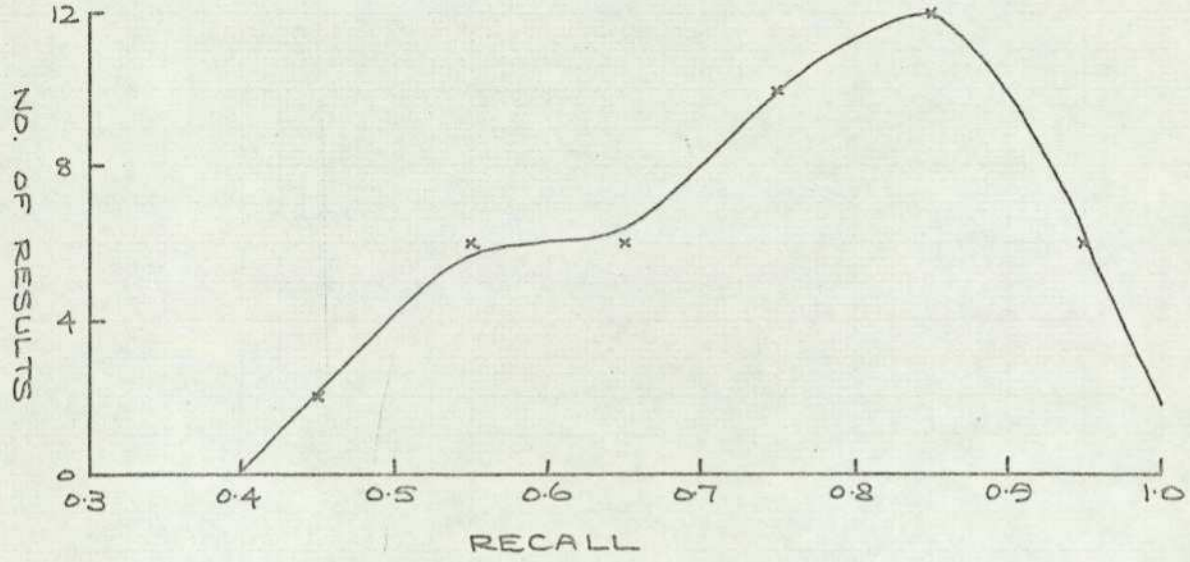


Recall frequency distribution (0·05 intervals) (set 712)

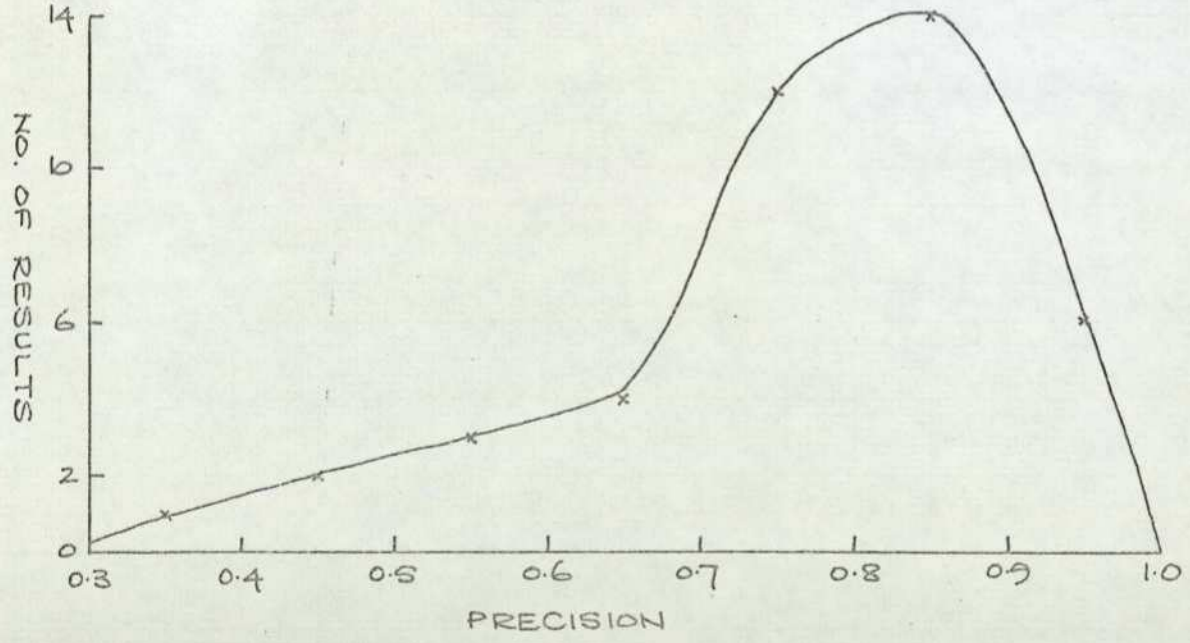


Precision frequency distribution (0·05 intervals) (set 712)

FIGURE 38

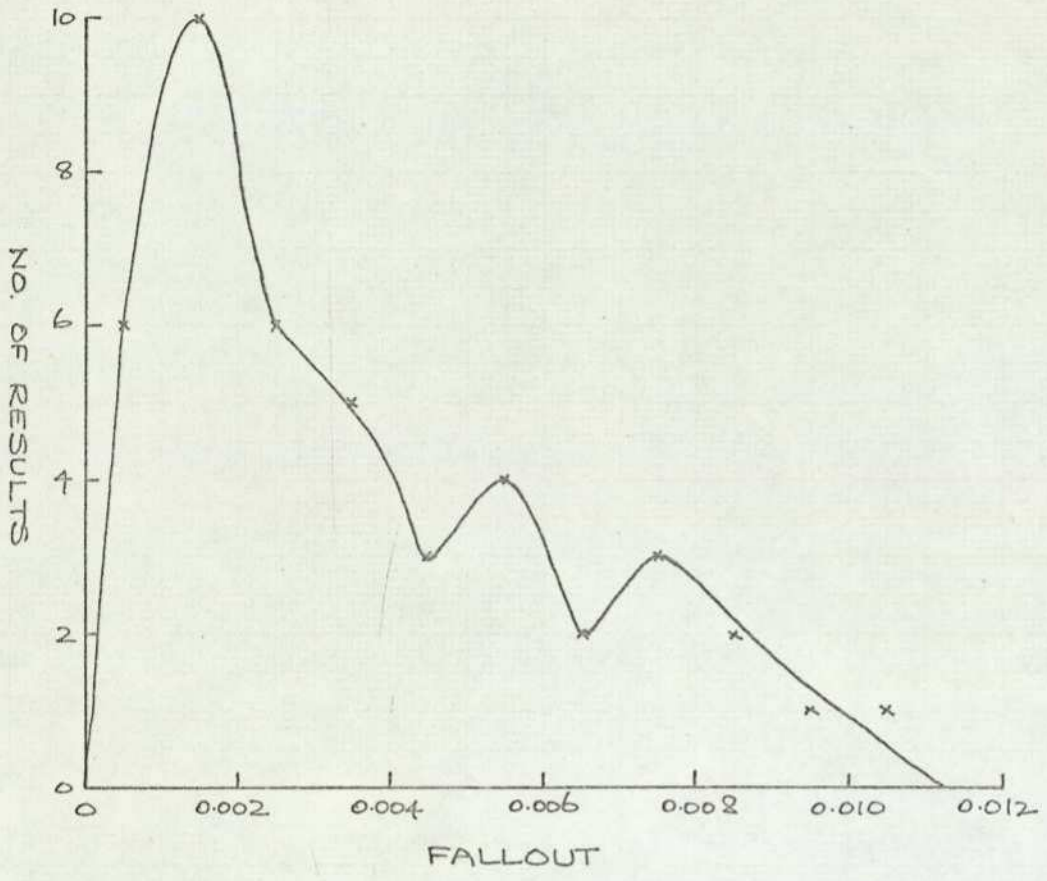


Recall frequency distribution (0·10 intervals) (set 712)

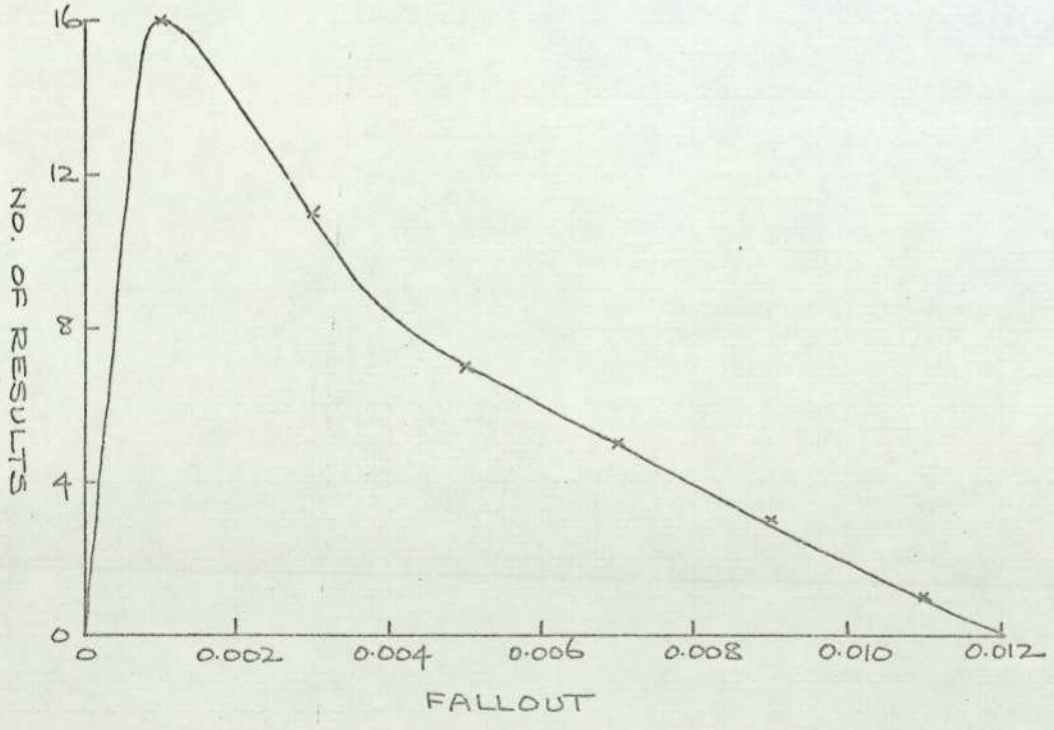


Precision frequency distribution (0·10 intervals) (set 712)

FIGURE 39

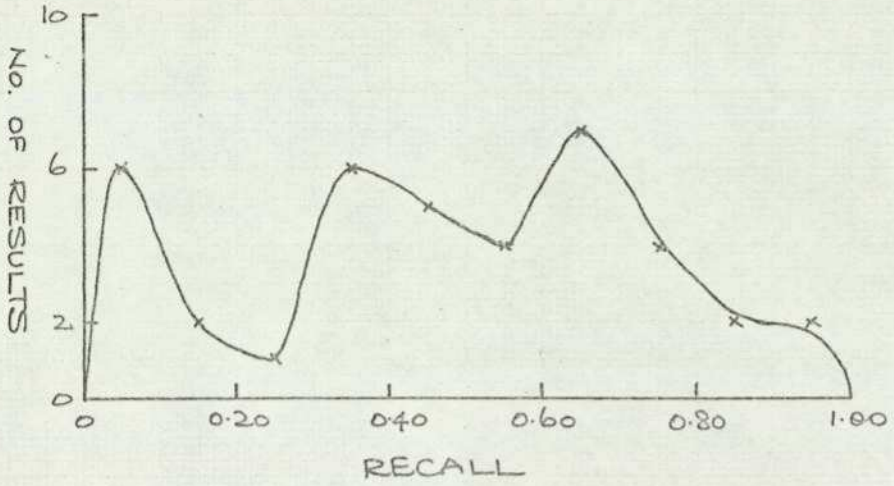


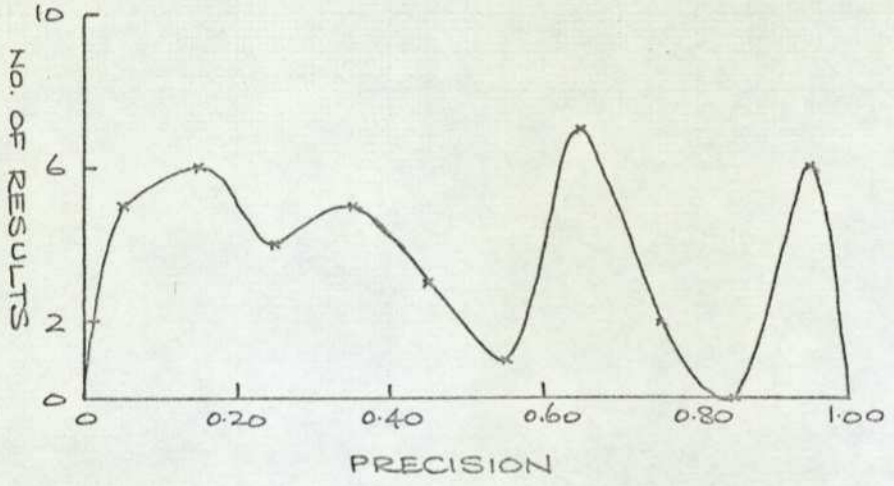
Fallout frequency distribution (0.001 intervals) (set 712)

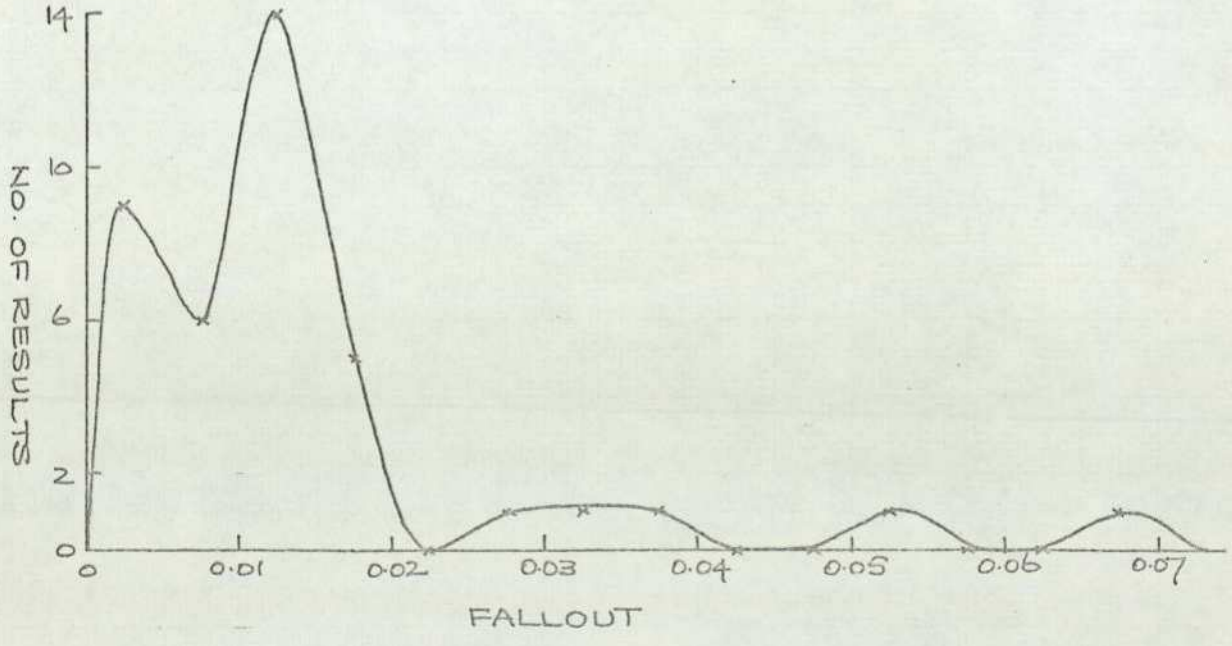


Fallout frequency distribution (0.002 intervals) (set 712)

FIGURE 40

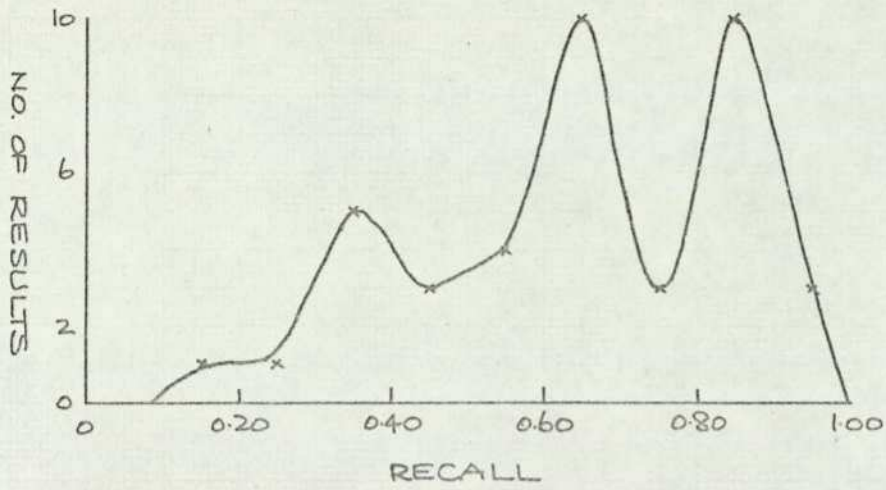


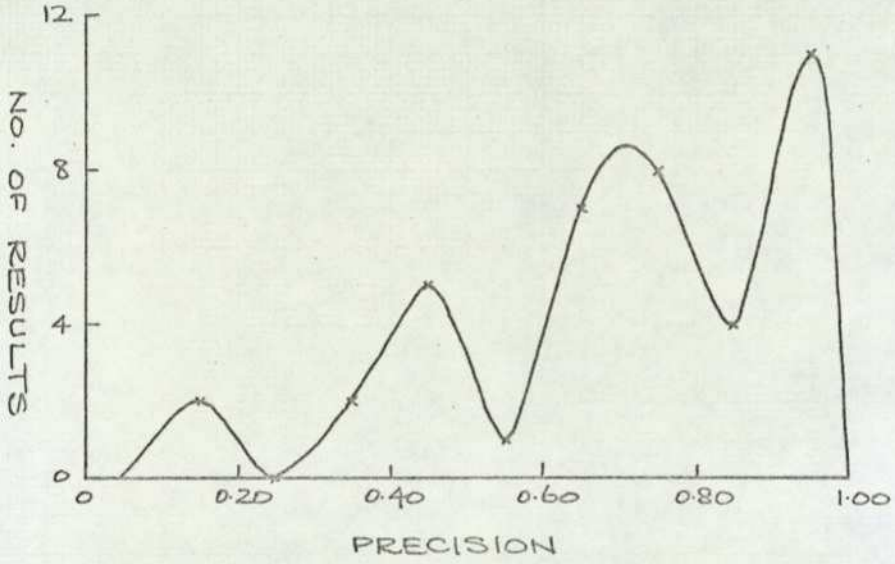


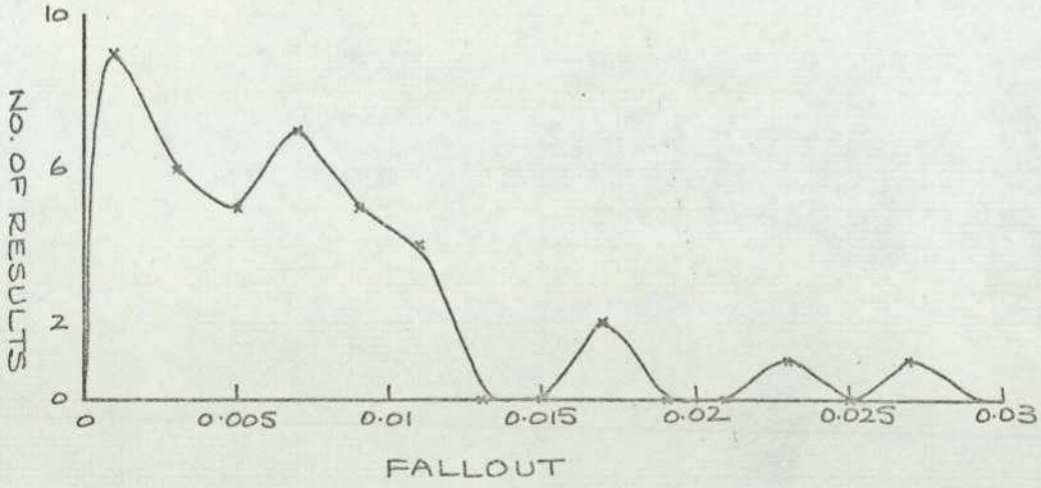


Recall, precision and fallout distributions (set 121)

FIGURE 41

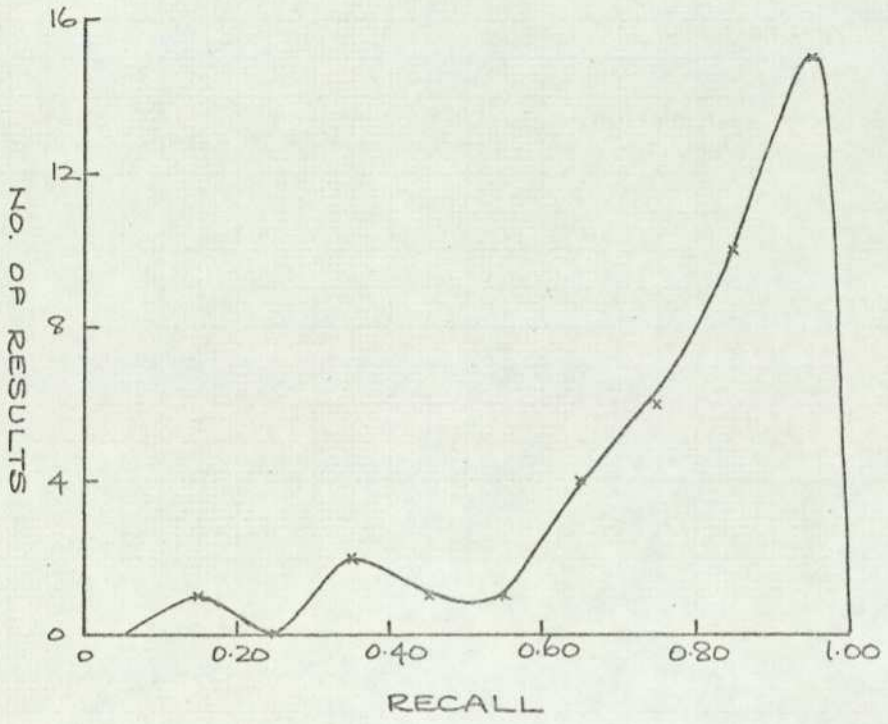


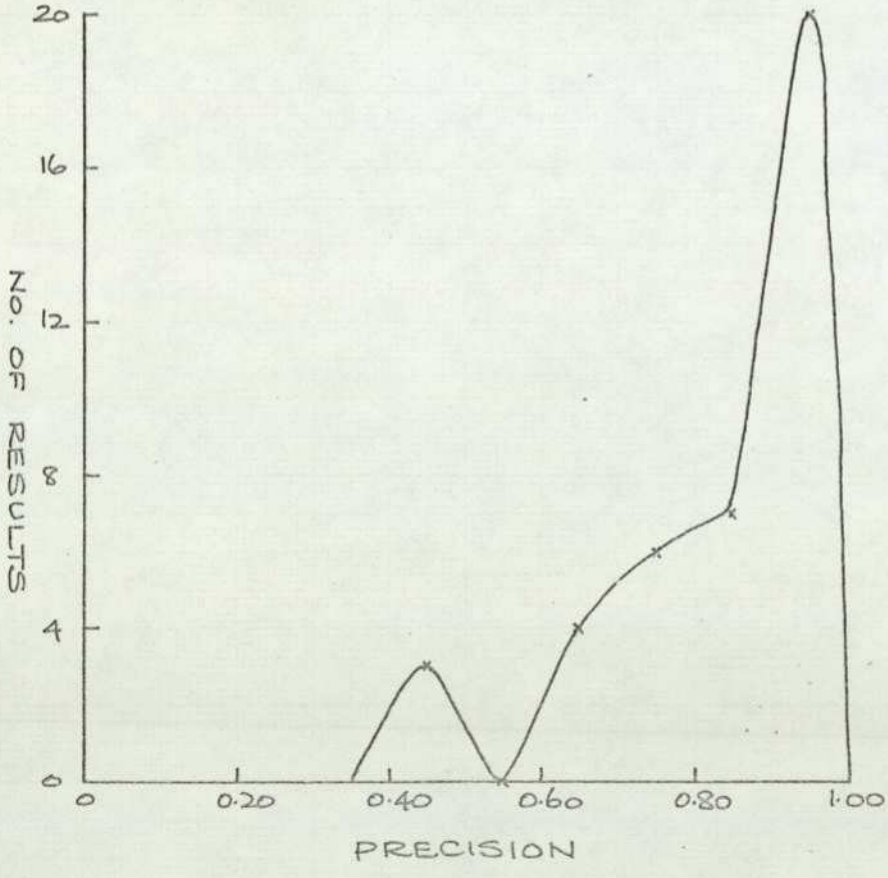




Recall, precision and fallout distributions (set 122)

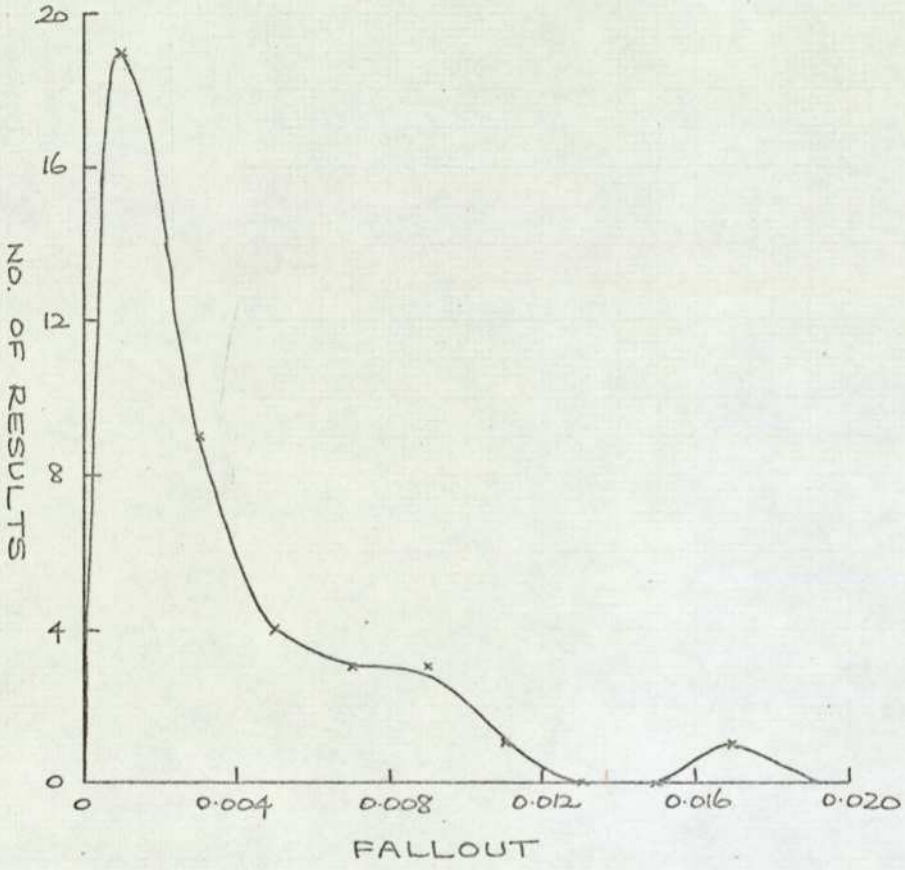
FIGURE 42





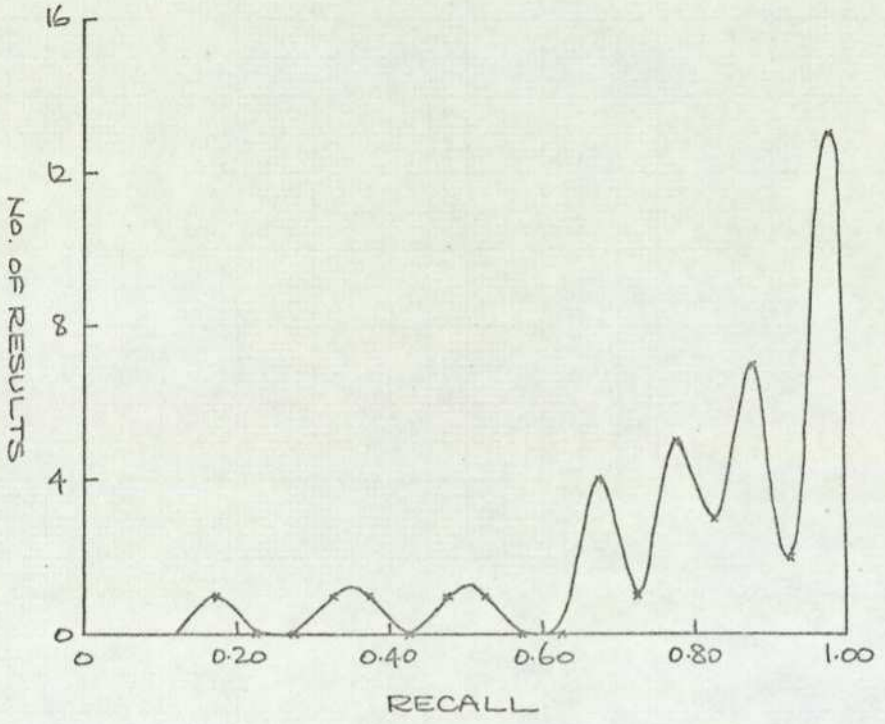
Recall, precision and fallout distributions (set 123)

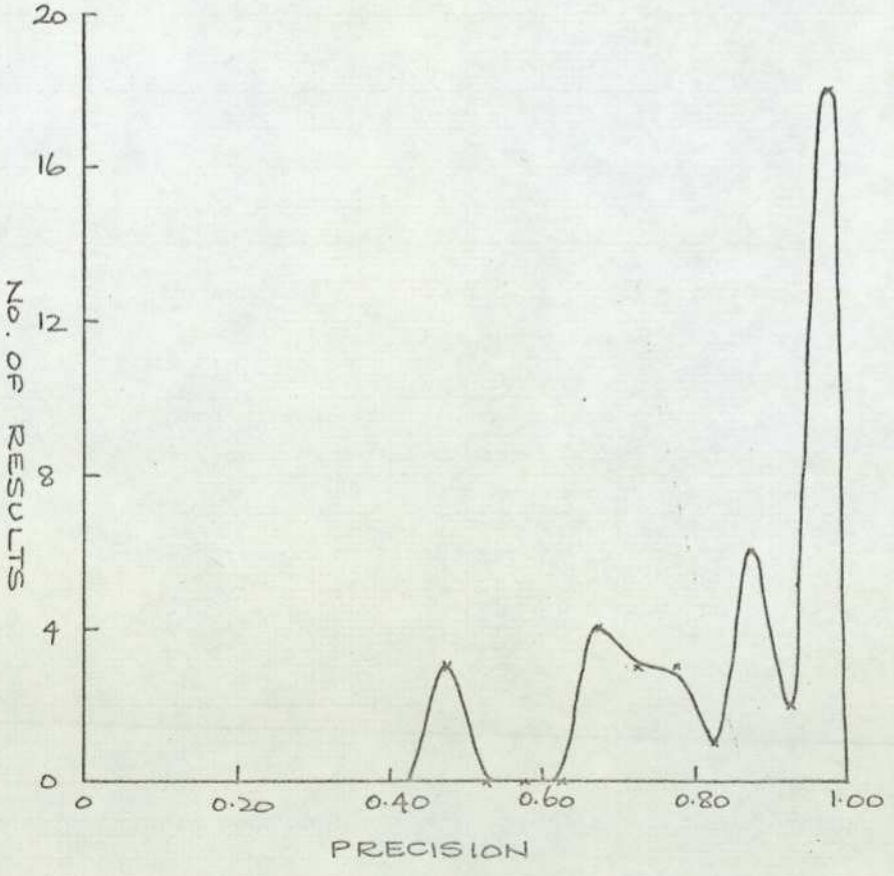
FIGURE 43



Recall, precision and fallout distributions (set 123)

FIGURE 43 (continued)





Recall and precision distributions (0·05 intervals) (set 123)

FIGURE 44

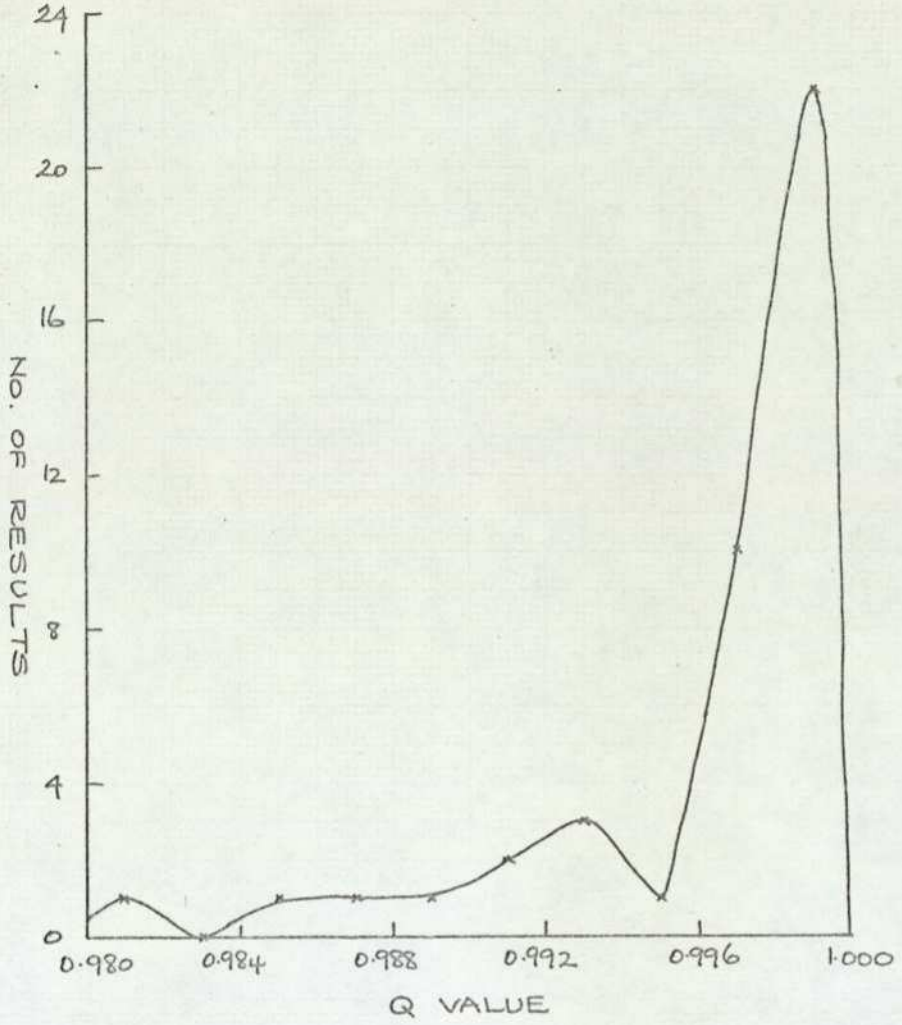


FIGURE 45 Q value distribution (set 712)

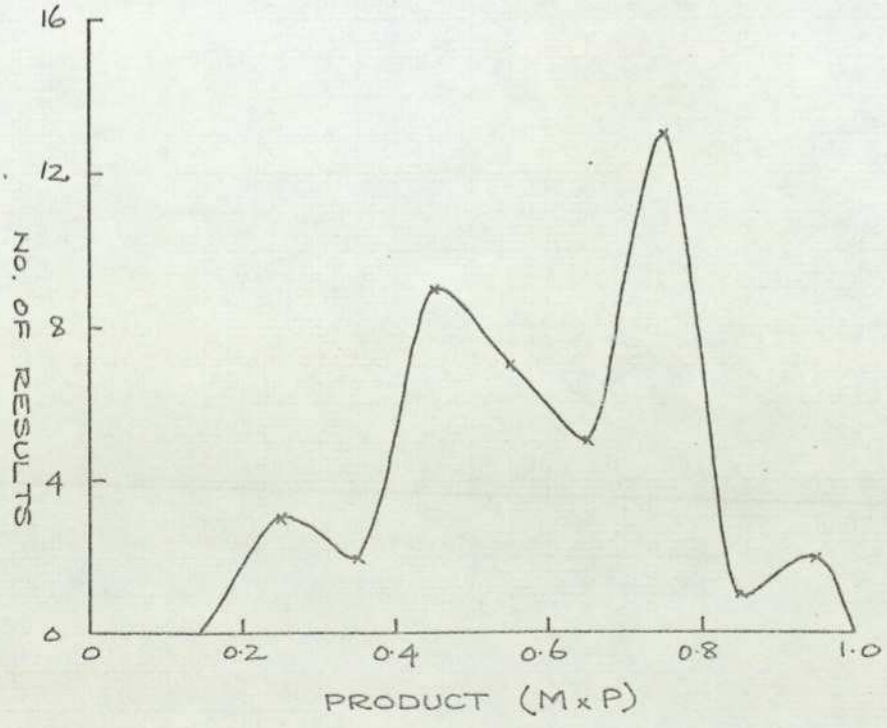


FIGURE 46 (M x P) distribution (set 712)

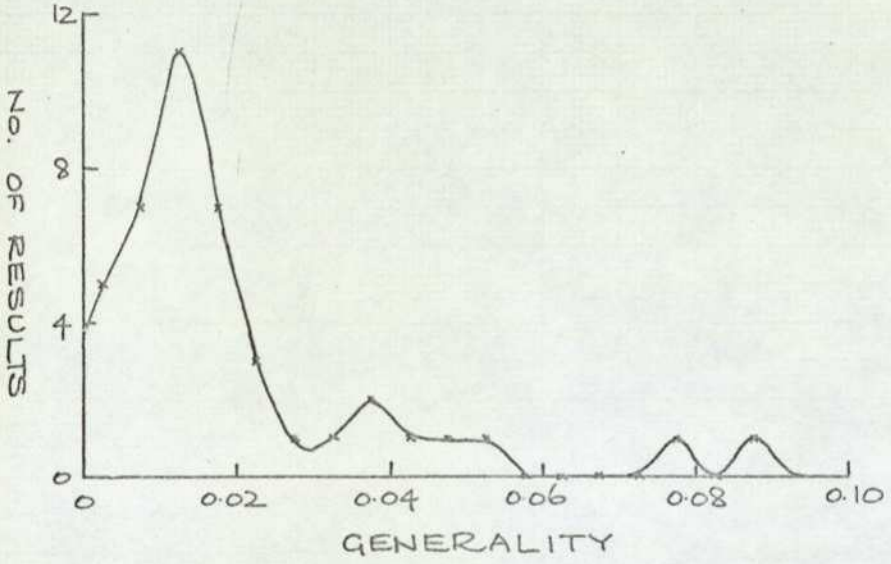
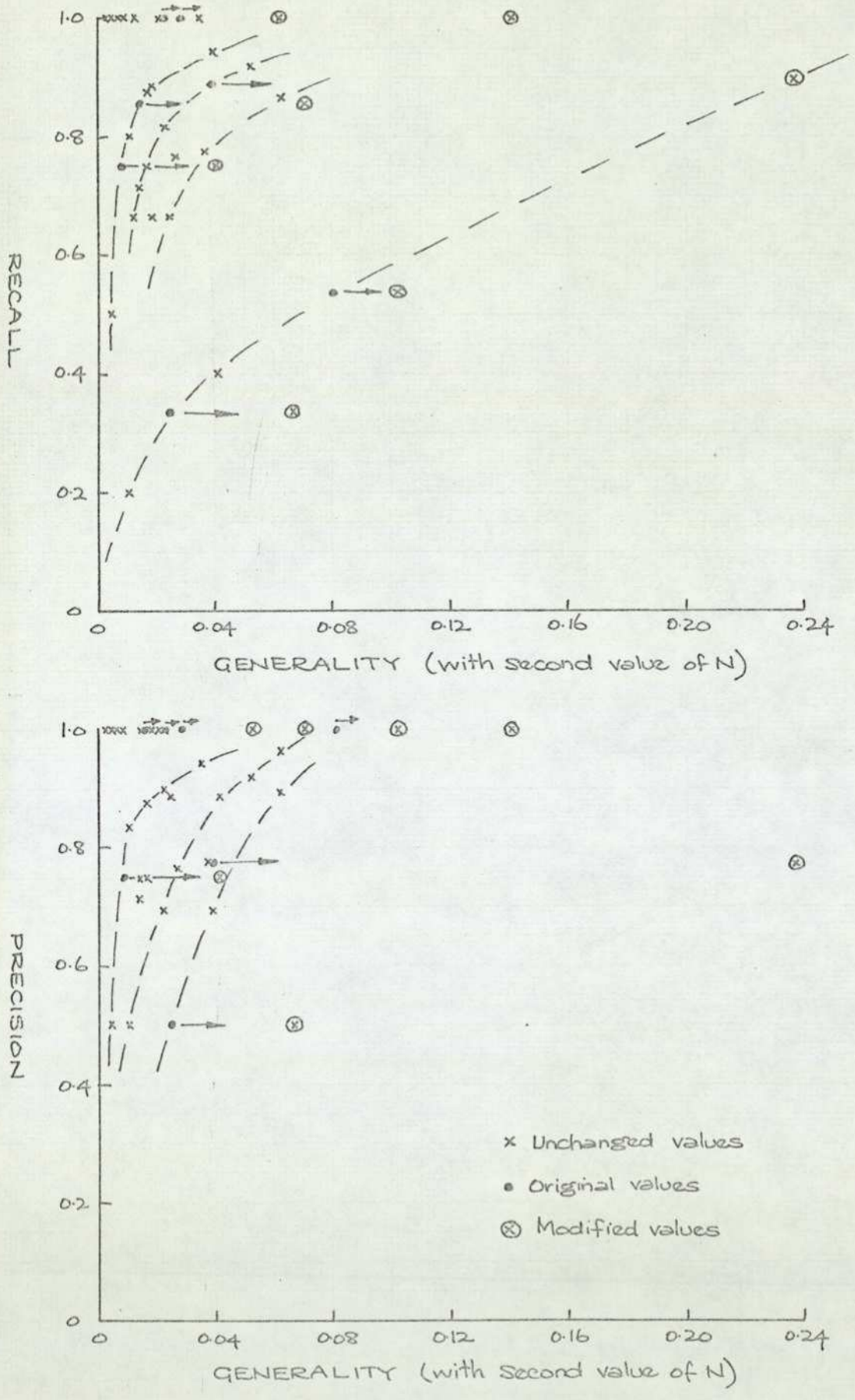
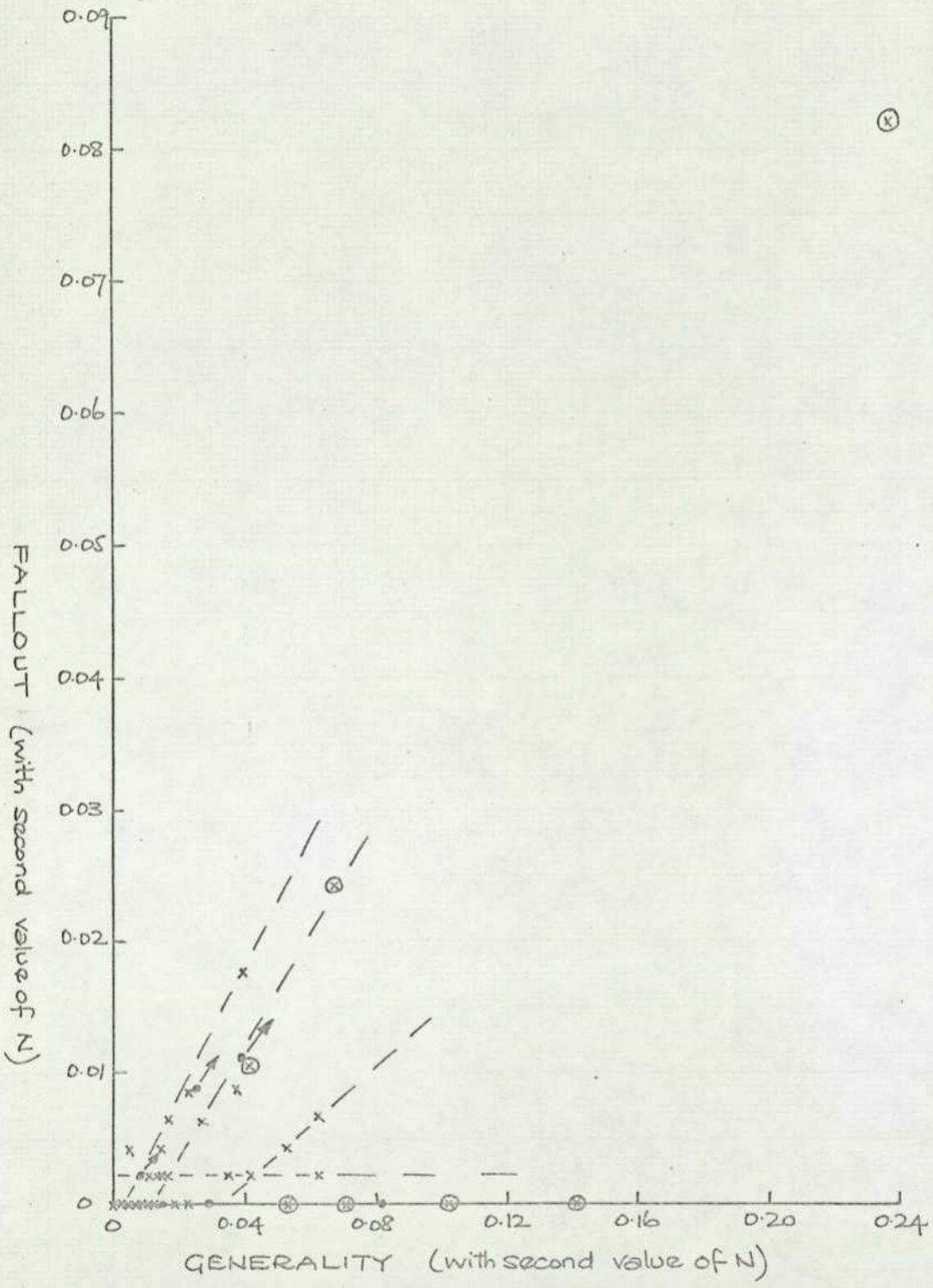


FIGURE 47 Generality distribution (set 712)



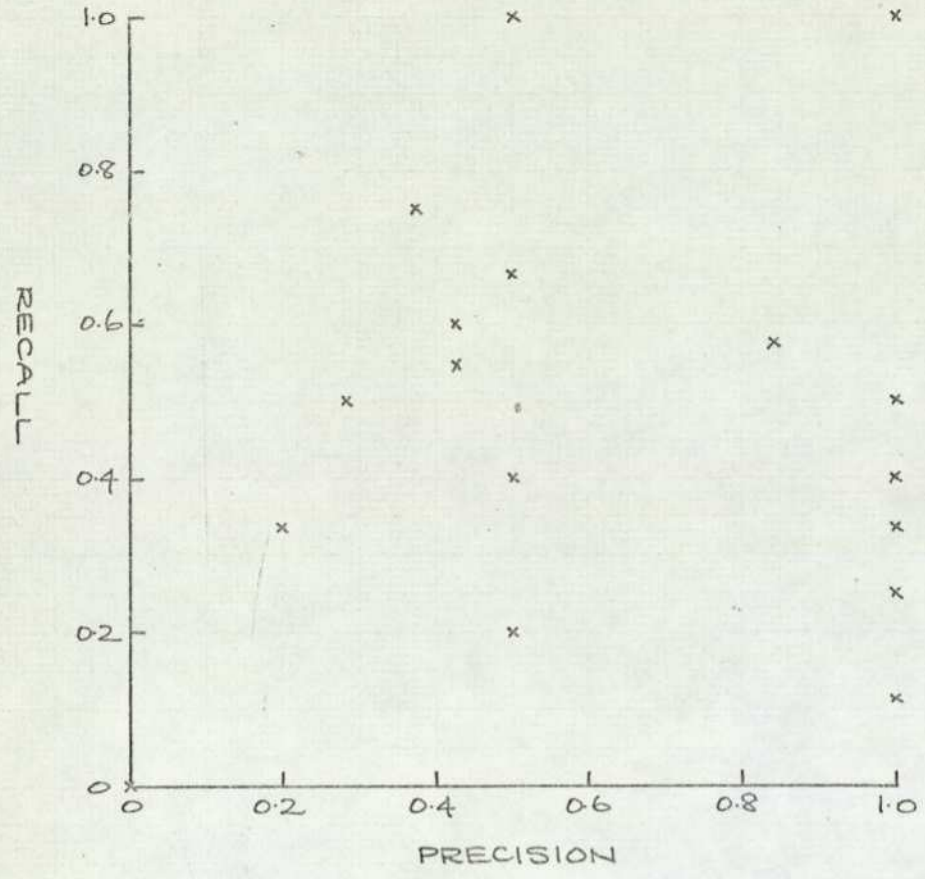
Recall - generality and precision - generality graphs (set-123) with second value of N

FIGURE 48

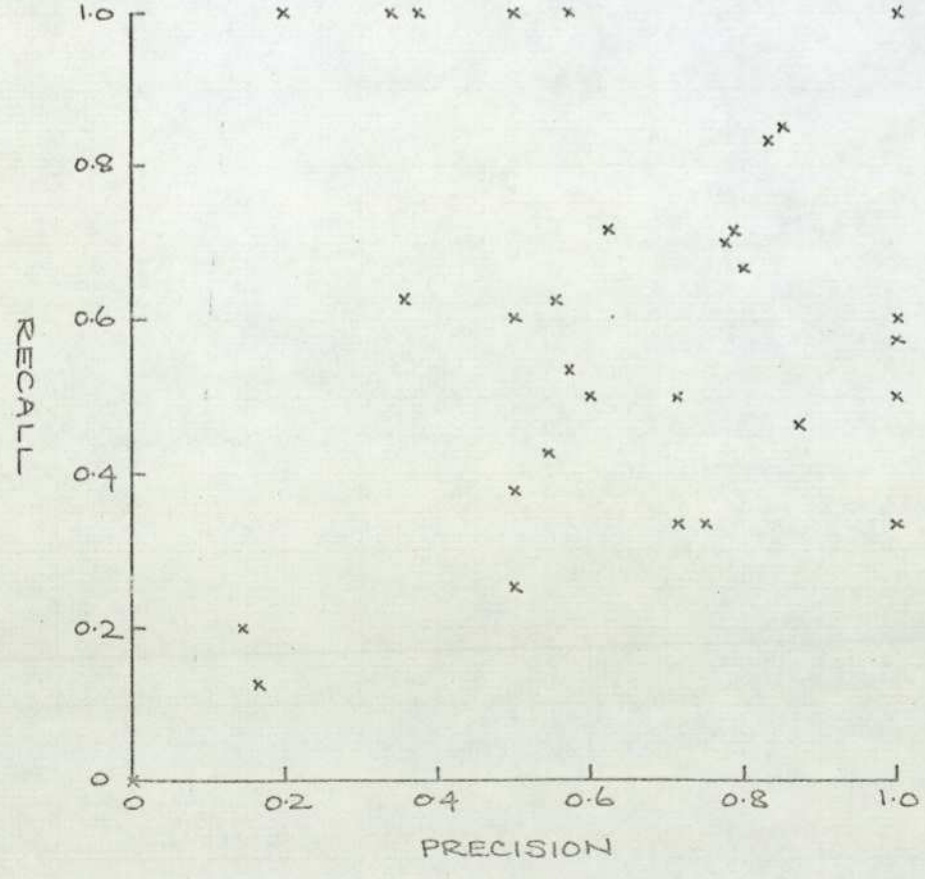


Fallout - generality graph (set-123)
with second value of N.

FIGURE 49

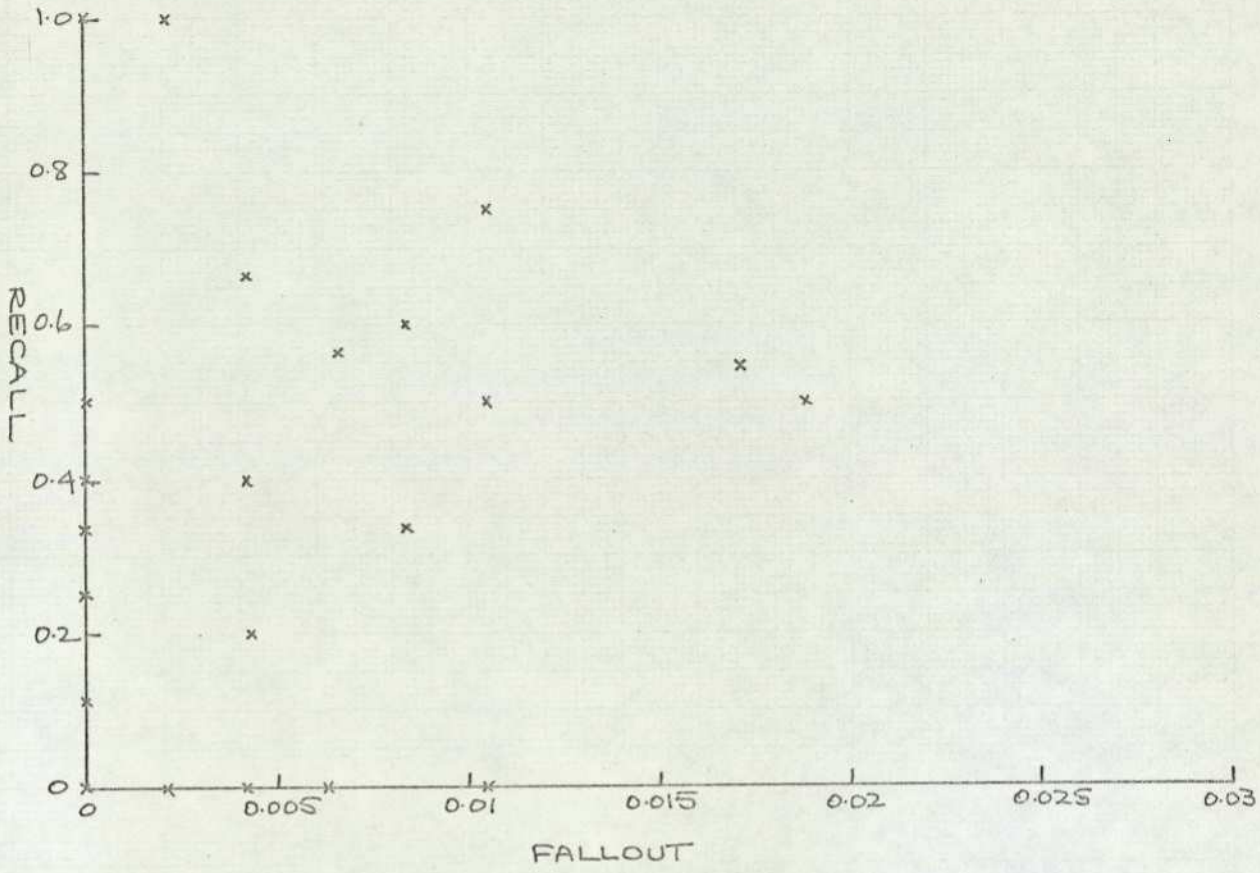


Recall – precision diagram (set 123, relevance 1)

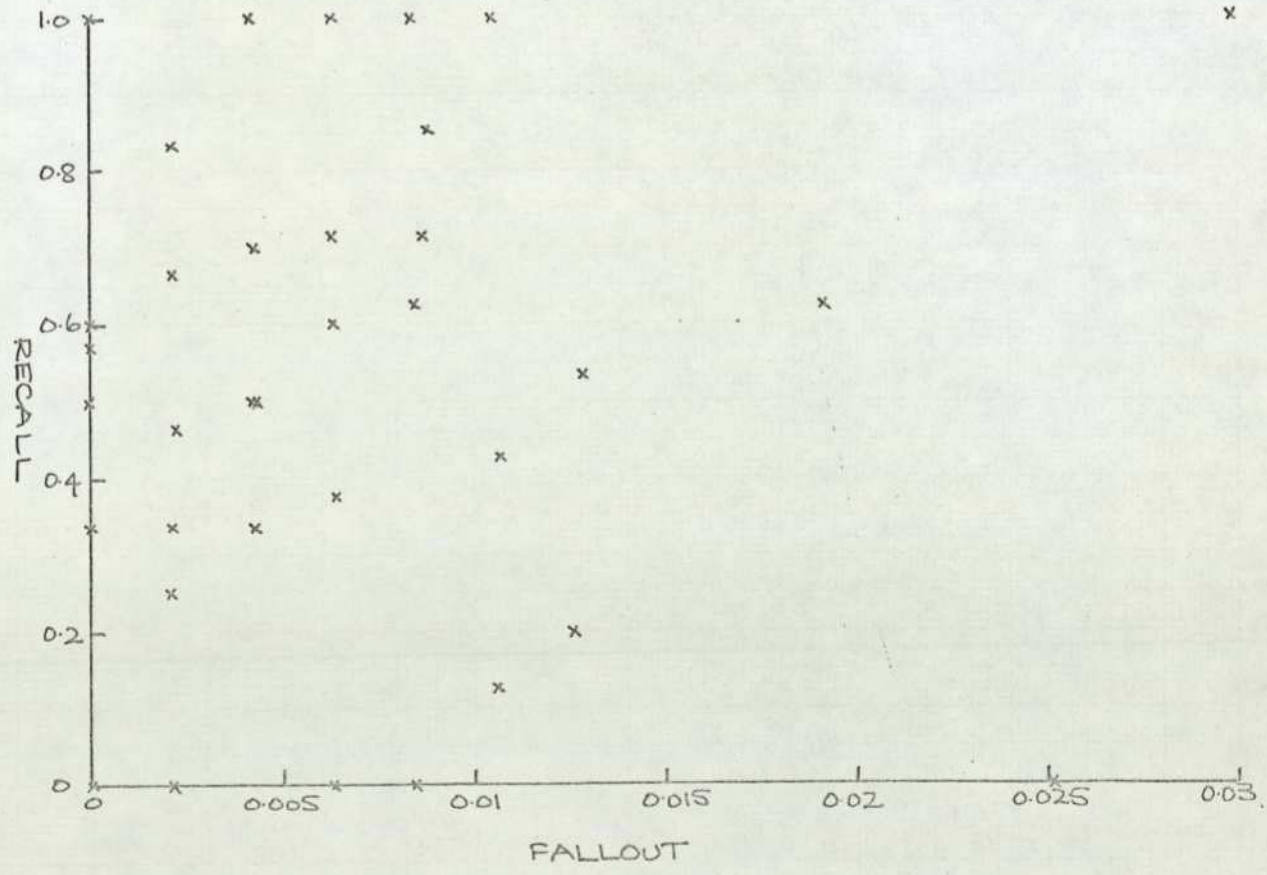


Recall – precision diagram (set 123, relevance 2)

FIGURE 50



Recall - fallout diagram (set 123, relevance 1)



Recall - fallout diagram (set 123, relevance 2)

FIGURE 51

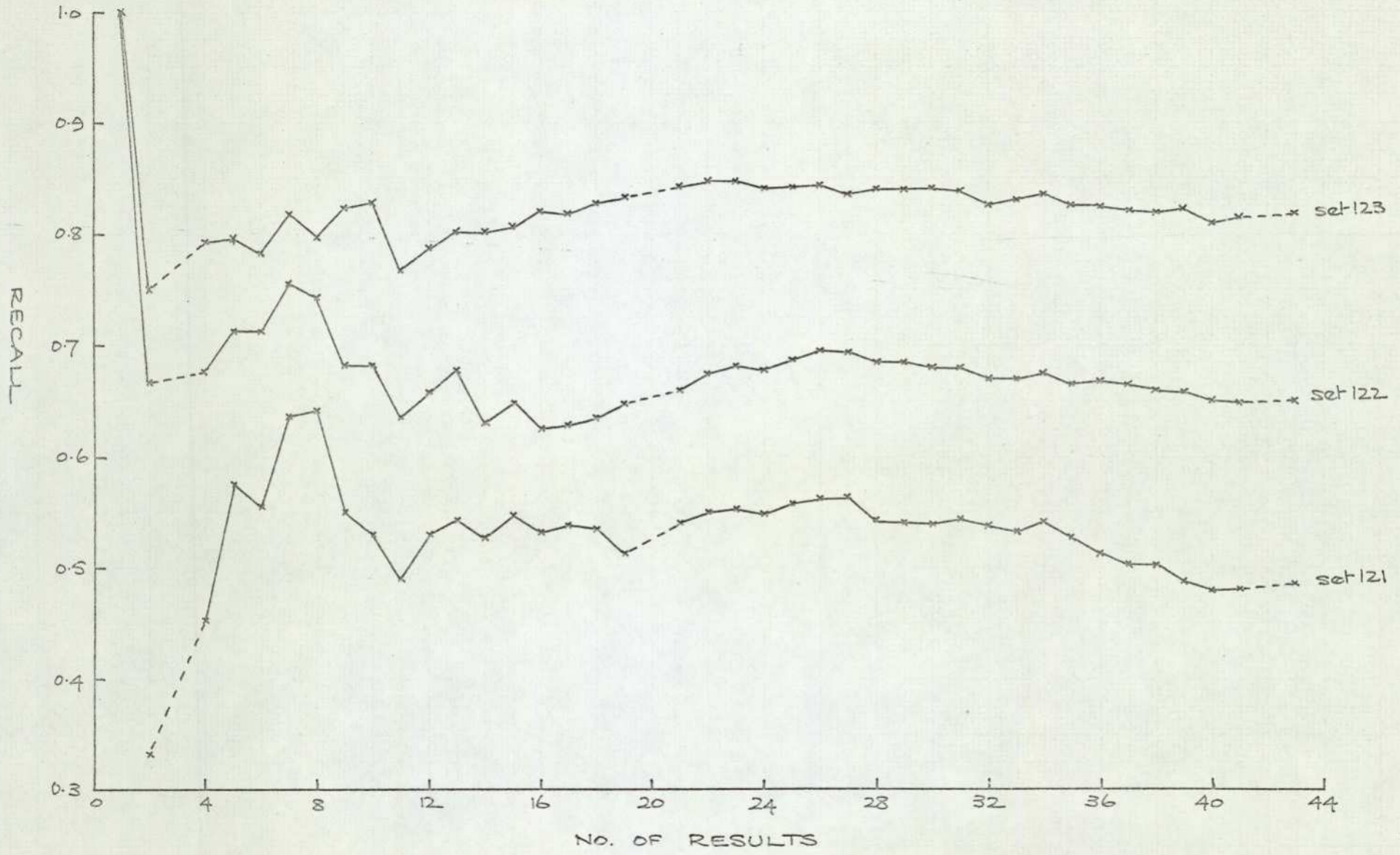


FIGURE 52 Cumulative average of recall (sets 121, 122 and 123)

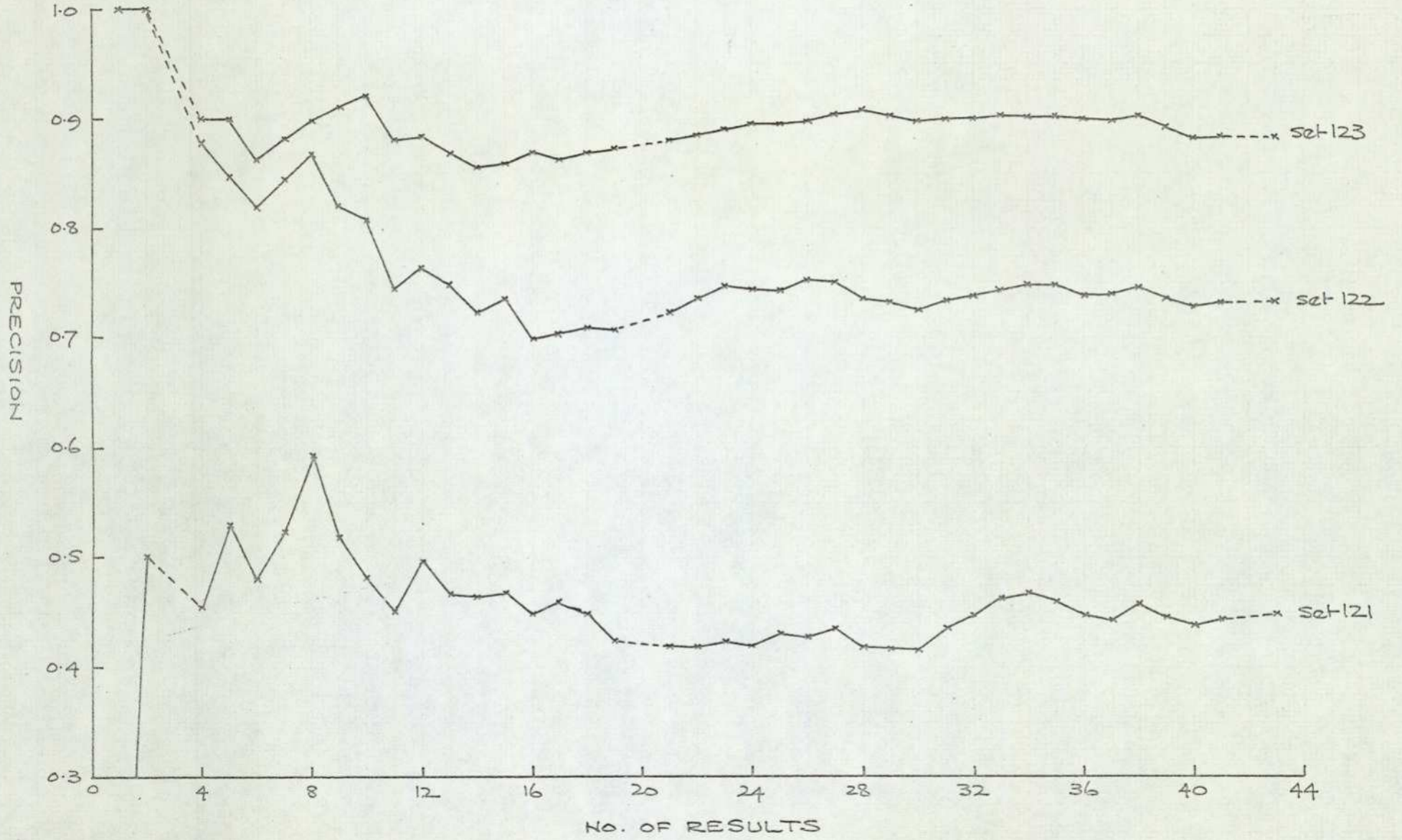


FIGURE 53 Cumulative average of precision (sets 121, 122 and 123)

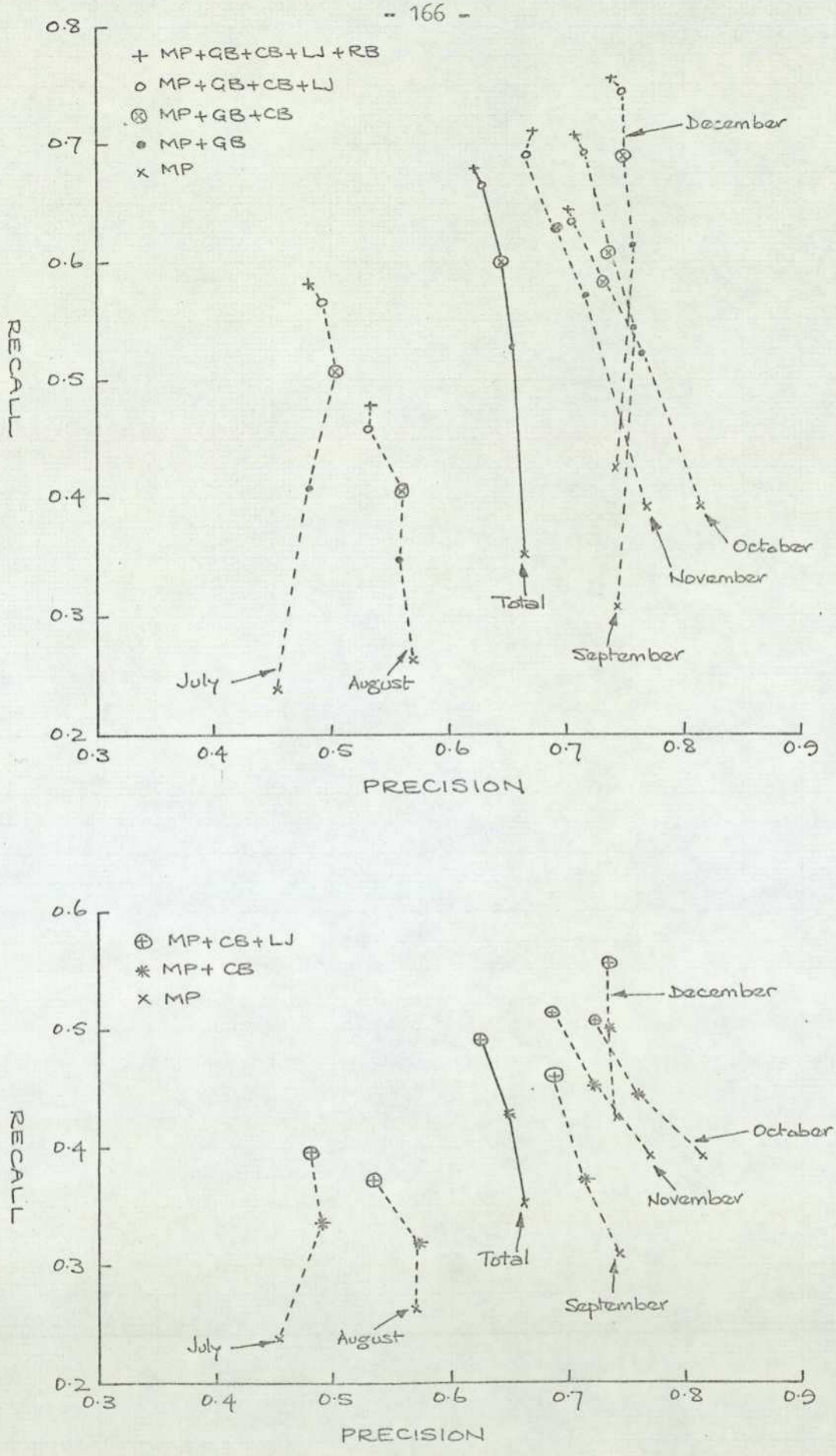


FIGURE 54 Recall-precision diagrams for selection strategies

3.3 S D I Systems

One of the main differences between SDI and retrospective search services lies in the profile. Although at first sight there may appear to be little distinction between the question posed to a retrospective search system and the SDI profile, inevitably the profile has a much greater and lasting impact inasmuch as it may be considered a repetitive question, and its properties will affect not just one search, but every subsequent search conducted in relation to it. The properties of SDI profiles are therefore worthy of more detailed consideration.

3.3.1 Profile collection and construction

The collection of profiles, as described in Section 2.4, was achieved by personal discussion or, more frequently, by correspondence with the user. It seemed surprising during the experimental part of the project that the interest statements which had been obtained by personal discussion did not lead to profiles which in all cases performed better than those obtained from letters. This was attributed to the fact that possibly the discipline imposed by committing one's thoughts to paper forced a situation where the user had to think more deeply and more clearly about his actual information need, whereas a discussion allowed a more rambling, incoherent and undisciplined attitude, particularly if the user had not given prior consideration to the question. The user tended to forget some aspect or detail, but returned to it later when perhaps some other aspect jogged

his memory. A letter should not permit such incoherence and confused thinking.

Of the 43 users who completed the project, the profiles of only eight were obtained by personal discussion. The averages of the results obtained for these profiles in set 712 were:

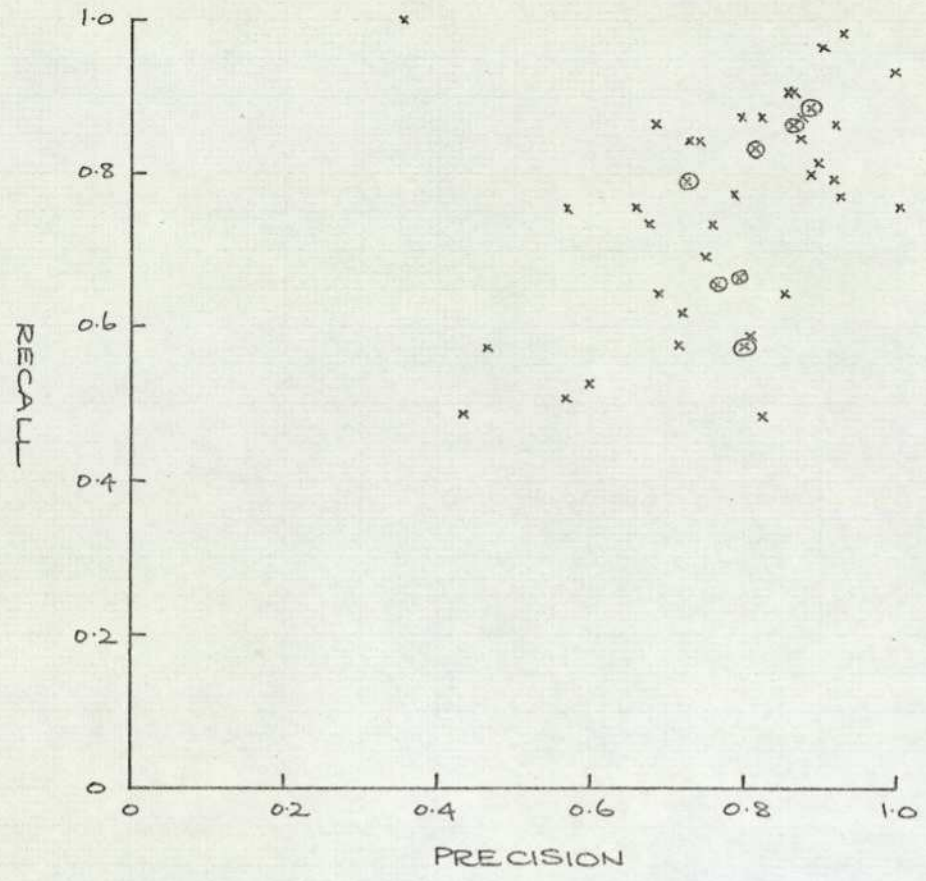
recall = 0.754, precision = 0.807 and fallout = 0.0030

compared with the overall averages for that set of

recall = 0.759, precision = 0.774 and fallout = 0.0037.

The individual results (circled) can be seen more clearly in Fig. 55.

In the light of the results obtained in the MEDLARS evaluation, these effects are not so surprising. There the best performance results were obtained when the user sent his request, written in his own natural language, by post to a MEDLARS centre. In that way the influence of the system, from both the logical and linguistic viewpoints, is kept to a minimum. As a result of this finding, a search request form was re-designed [62] which, although requiring effort by the user, endeavours to extract from him, in a formalised manner, a statement which accurately describes his true information need. It also encourages him to specify parameters and details which are in his mind at the time of compiling the request, but escape verbalization.



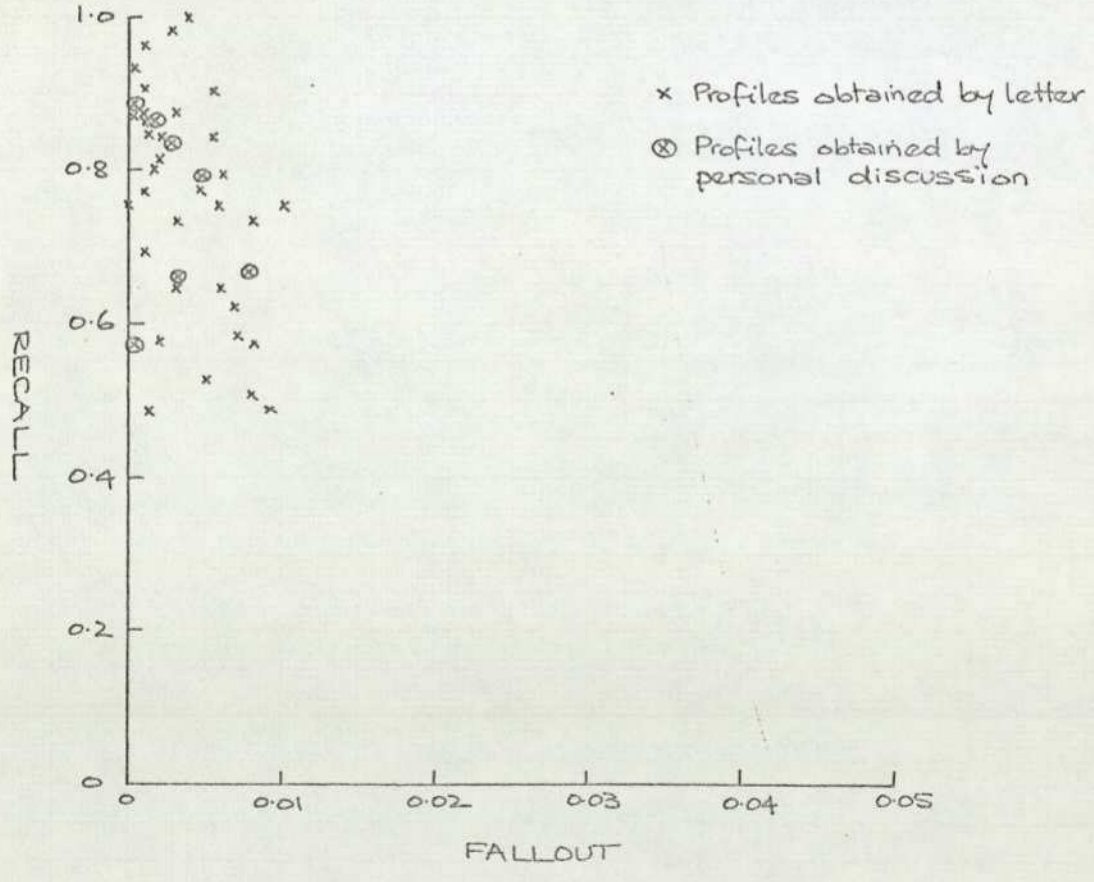


FIGURE 55 Recall – precision and recall – fallout diagrams showing different profile interaction (set 712)

The results from the INSPEC evaluation [13] are in partial agreement with those from MEDLARS. Their interest statements were initially obtained in writing and they investigated various types of further user - system interaction, namely no interaction; submission of the profile to the user for comments; questions asked of the user if necessary; interview of the user; and interaction as required. The conclusion was that there was a decided benefit in interaction, although there was no marked advantage in any one particular form of interaction.

Although contrary to the service concept, it may prove necessary in profile collection and construction to demand greater effort and active involvement by the user, as in the case of the MEDLARS search request form. One benefit of more active involvement early in the search process may be deeper interest and a greater appreciation of the problems and the results that may be encountered later. Although there are no comparative statistics to support such a theory, user involvement and participation has always been part of the UKCIS policy [97]. The UKCIS user has to construct his own profile, although he may receive instruction from UKCIS personnel, as well as aids such as the Search Manual and the KLIC index. In their system, the influences, both logical and linguistic, are far from minimised, but considerable involvement and careful thinking about the actual requirements and information need are necessary to obtain satisfactory results.

However, with regard to profile modification the situation may be slightly different. In the present research, the

majority of interaction was by letter, mainly because it proved time-consuming and frequently unsuccessful to contact users by telephone. But it was noticed, in many instances when telephone contact was made, that users expressed a preference for such conversations because they felt that problems were easier to explain and discuss in the more direct interaction provided by the telephone than by letter. A similar effect was noted by Scheffler and March [89] who found direct contact with users effective in obtaining uninhibited feedback and meaningful expressions of the users' interests.

3.3.2 Analysis of profiles

The amount of literature reporting a detailed analysis of profiles seems to be limited. Housman's survey [54] includes a broad analysis of the profiles of the one hundred systems investigated. The analysis considers, among other things, the number of terms per profile in operational and pilot SDI systems and shows two peaks, one at 10 - 15 terms per profile (12 systems) and the other at 25 - 50 terms per profile (13 systems).

Profile structures were also considered in some detail in the UKCIS evaluation [7]. A typical run, of the main experiment, had an average of 23.3 terms per profile, which does not fit into either of the peaks of Housman's survey. They also found between 1 and 3 parameters per profile, the higher number of parameters being associated with weighted profiles. Other analyses related

to term types, use of NOT logic and weighting, logic type and truncation modes. Housman's survey analysed profile preparation and coding of profile terms. Such analyses have little relevance in the present context.

It was felt that the profiles in the present project were basically simple in structure, although their size gave a superficial impression of complexity. It was also obvious that most profiles could be subdivided into several sections. The profiles of the 43 participants who completed the project were analysed. For this analysis some commonly used terminology was defined. In other work the lack of definition of some terms has led to difficulty in comparing results. A "profile" was interpreted as the interests of one user or participant. Thus the number of profiles is equal to the number of participants. The profiles were subdivided into "sections" by the factors that the parts have in common. A new section was prepared when there was a significant change in the main aspect of the profile. For example, "the saturation magnetization or electrical conductivity of Ni-Mn alloys" would be regarded as one section, but if "the relation between atomic order and magnetic properties of Ni alloys" was also required, this would be treated as a separate section, even though there are certain common features in the two sections. Thus, for the purposes of this research, a "section" can be defined as a separate part of the profile not related or not closely related to any other section.

A section is composed of one or more "parameters", linked by relational operators. Each parameter may contain a

number of "terms". In the previous example, "saturation magnetization" and "electrical conductivity" are two terms comprising one parameter, with "alloys" and "Ni-Mn" as two other parameters. By analogy with Boolean algebra terminology, parameters are linked with AND or NOT logic, whereas terms are usually linked by AND or OR logic. On this basis the 43 profiles were analysed.

The average number of sections per profile was 5.4. The variation was quite wide, with a few profiles having only one section and one profile having as many as 14 sections. Within sections an average of 3.56 parameters was found. The average number of terms per profile was 35.2, again with a wide variation from 2 to 108. This gave an average of 6.5 terms per section and 1.8 terms per parameter. Thus this "pilot system" would fall into the second peak (25 - 50 terms per profile) found in Housman's survey. The results are somewhat different from those found in the UKCIS evaluation [6, 7] and the comparative figures are given in Table 27. It should be noted that the UKCIS concept of a "profile" is not completely clear, but it is assumed that their "profile" is equivalent to a "section" and the figures have been quoted accordingly.

So far the analysis is entirely quantitative, but the structure of the sections, i.e. their two-dimensional display incorporating relational operators, is also relevant in this analysis. A one-parameter section can be represented by the term A and a two-parameter section can be similarly represented by A / B.

TABLE 27

Comparative profile analysis

	Present project	UKCIS
Sections/profile	5.4	2.4
Terms/profile	35.2	-
Parameters/section	3.6	1 - 3
Terms/section	6.5	23.3
Terms/parameter	1.8	-

With three-parameter sections there are more possibilities and different structures are discernible, e.g. A / B / C, A / B.
/
C

The sections can exist in basic structures, or combinations thereof, and as the number of parameters increases, so also do the possible combinations. The basic structures have been termed linear, two-dimensional linear, circular, and star. "Combined" is a combination of two (or more) of the other types. The basic structures for various numbers of parameters are indicated in Table 28.

There are various other forms which such structures can take. For example, two terms linked by a curly bracket ({}), indicating that the two terms must be present simultaneously, are regarded as structurally equivalent to one parameter and on this basis, five different combinations of four terms were found,

TABLE 28

Basic structures with varying numbers of parameters

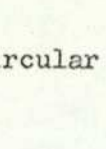
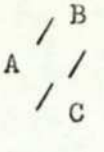
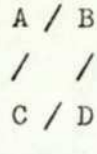
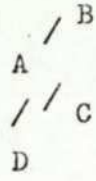
	Number of terms			
	3	4	5	6
Linear	A/B/C	A/B/C/D	A/B/C/D/E	A/B/C/D/E/F
Two-dimensional linear	A / B / C	A / B / C / D	A / B / C / D / E	A / B / C / D / E / F
Circular				
Star	-			

TABLE 29

Number of forms of each type of structure

	Number of terms							
	3	4	5	6	7	8	9	10
Linear (L)	3	5	5	1	-	-	-	-
Two-dimensional linear (LL)	2	4	9	3	4	1	-	-
Circular (C)	2	2	1	1	-	-	-	-
Star (S)	-	1	1	-	-	-	-	-
Combined	-	5(C+L)	2(C+L)	3(C+L)	1(C+L)	-	-	-
					1(C+LL)			
			1(S+L)	1(S+L)	1(S+L)			1(S+LL)

which were classified as linear, viz.

A/B/C/D $A/\begin{Bmatrix}A\\B\end{Bmatrix}/C/D$ $A/\begin{Bmatrix}B\\C\end{Bmatrix}/D$ $A/\begin{Bmatrix}B\\C/D\end{Bmatrix}$ and $A/\begin{Bmatrix}B\\C\\D\end{Bmatrix}$

A table showing the number of different forms of each structure is given in Table 29, from which it can be seen that the number of different forms of each structure approximately increases up to 5 terms and then falls off rapidly (due to the few sections having 6 or more parameters).

To demonstrate more clearly the more typical section structures Table 30 shows the number of occurrences of each type of structure. As would be inferred from the average number of parameters per section (3.6) the highest frequencies occur with 2, 3 or 4 terms.

TABLE 30

Number of occurrences of each type of structure

	Number of terms										
	1	2	3	4	5	6	7	8	9	10	Total
Linear	7	42	39	28	9	3	-	-	-	-	128
Two-dimensional linear	-	-	40	24	19	2	3	1	-	-	89
Circular	-	-	6	3	2	-	-	-	-	-	11
Star	-	-	-	4	1	1	-	-	-	-	6
Combined	-	-	-	6 C+L	2 C+I	3 C+L	1 C+L	-	-	-	
							1 C+LL				
					2 S+L	2 S+L	1 S+L			1 S+LL	19
Total	7	42	85	65	35	11	6	1	-	1	253

Accordingly it is demonstrated that the most frequently used types of structures are the linear and two-dimensional linear.

The tables and the quantitative analysis show that the average profile in this project has 5.4 sections, each of which has 3.6 parameters and a total of 35.2 terms per profile, which are arranged in a linear or two-dimensional linear display.

Altogether there are 253 sections of profiles, applying to the 43 users. Thus 253 questions were posed to each of the six issues of Metals Abstracts. Unfortunately the results could not be analysed in terms of these separate questions for several reasons. Each user provided a list of relevant abstracts for his total profile and, in very many cases, it was not possible to accurately assign the abstracts selected to one particular section of the profile. Some abstracts would of course be applicable to more than one section of a profile. It proved difficult to obtain full co-operation from the users in relation to the selection and non-selection of abstracts and it was considered that attempting to assign abstracts to sections would have increased both the difficulties and the inaccuracies associated with the results. A correlation between profile structure and results also proved to be an unrealistic proposition. Each profile contained several sections, with no consistency between sections of either number of parameters or basic structure. Thus there was no indication that one user's separate interests are structurally similar. In future research however, it would be interesting to investigate more fully the performance of different structures in relation to specific

user information needs and thus, perhaps, the information scientist/system designer would be better able to advise user/clients on a suitable structure to meet those needs.

3.3.3 Mode of operation

In Section 3.2.8 (Distribution of results) the multi-modal distributions of the performance measures were discussed and in general it seemed as though there were three or four modes for each measure. This characteristic was investigated more fully and it was found that the individual results did not necessarily fall into the same mode for the various measures.

By comparing the recall and precision results it is possible to obtain "pure modes", i.e. profiles which fall into the same mode for each parameter. The result is as follows:

Mode I	Profiles 3, 12, 21, 24, 32, 35, 47, 56
Intermediate	42
Mode II	7, 9, 10, 14, 26, 66, 69
Mode III	13, 57, 64, 65

Mode I has the highest group of recall and precision results and Mode III has the lowest results. A recall - fallout comparison yields similar results, except that the components of the groups are larger. These groups are:

Mode I	Profiles 3, 12, 14, 21, 24, 32, 35, 37, 42, 43, 47, 51, 56, 59, 61, 67

Mode II Profiles 7, 10, 29, 66, 69

Mode III 26, 27, 34, 57, 60, 64, 65

Finally, pure modes can be obtained for profiles which fall into the same modal group for each performance measure and these are:

Mode I Profiles 3, 12, 21, 24, 32, 35, 47, 56

Mode II 7, 10, 66, 69

Mode III 57, 64, 65

For recall - precision comparisons, 20 profiles fall into pure modal groups; for the recall - fallout comparison, 28 profiles fall into pure modal groups, and for recall - precision - fallout comparisons, only 15 profiles fall into the pure groups. The remaining 28 profiles fall into different groups for the various performance measures, i.e. they may have high recall and low fallout and precision, etc.

One of the more interesting aspects of these results is to consider why different profiles should produce varying results. For what reasons does one profile produce good recall, precision and fallout and another profile produce inferior recall, precision and fallout? In a test of this kind there are many possible reasons which may be associated with the many human factors involved in the system. However, it is reasonable to suppose that the human factors associated with the searching operations would have been reasonably constant for all profiles. Human factors associated with the user will not of course be constant, and to some extent these have been catered for. The most likely reason for differences of these types would seem to lie with

the profile and this area has been studied in greater detail for the pure modal groups from recall - precision comparison.

In comparing profiles there are various aspects that can be considered, some quantitative and others qualitative. The quantitative aspects include the number of sections per profile and the number of concepts per profile. The structure of the profile sections can also be considered and the analysis into linear, two-dimensional linear, circular, star and combinations is similar to that given in the previous section. A purely qualitative aspect is the author's own assessment of the simplicity or complexity of the profiles, according to a scale 1 to 4 allocated during the experimental part of the project. The interpretations of the values are: 1 - simple, 2 - relatively simple, 3 - average, 4 - complex. These four aspects are shown in Table 31. This tabulation indicates that the profiles in Mode I have different characteristics from those in Mode II and III but the latter two categories are closely similar. Thus profiles in Mode I have on average few sections and fewer concepts. According to the researcher's assessment they are generally simpler, but structurally the sections are not dissimilar. Thus one might conclude that "smaller" profiles perform better, but this does not account for the second and third modal distinctions.

A further analysis is to consider the actual concepts of which the profile is composed. In previous work [39] it has been suggested that this may affect performance. Consequently the profile sections were analysed according to whether the concepts

TABLE 31

Analysis of profiles by mode of performance

Profile	No. of sections	No. of concepts	Researcher's assessment	Structure
Mode I				
12	1	46	1	L
24	1	2	1	L
3	4	25	2	1/L/LL
21	3	27	2	L/L/LL
32	4	22	2	LL/LL/LL/L+C
35	3	24	2	L/L/LL
56	4	19	2/3	2/L/L/LL
47	4	41	2/3	L/L/LL/LL
Average	3.5	26	1.9	(1/2/11L/9LL/L+C)/8
Intermediate				
42	6	40	3	1/L/C/L+S/LL/LL
Mode II				
7	7	52	2/3	2/2/L/LL/LL/LL/LL
10	6	38	3/4	2/L/LL/LL/LL/LL
69	1	10	2/3	LL
14	7	38	2	L/L/L/LL/LL/LL/S
66	10	45	2/3	1/2/2/2/2/2/L/LL/LL/LL
26	5	73	4	1/2/L/LL/LL
9	8	48	4	2/2/2/L/LL/LL/LL/LL
Average	6.3	43	3.0	(2x1/12x2/8L/21LL/S)/7
Mode III				
64	5	29	2/3	2/L/C/S/LL
13	8	39	3	2/2/L/L/L/L/L/LL
57	8	54	3	2/L/L/L/L/L/L/LL
65	4	42	2/3	L/L/LL/C
Average	6.2	41	2.75	(4x2/14L/4LL/2C/S)/4

were entities, activities, abstracts, properties or heterogeneous terms. Also the main entity or material involved was analysed into chemical element, non-ferrous alloy, steel or compound. The results of this analysis are shown in Table 32 and this indicates that the type of material was not particularly significant except in so far as Mode I profiles were concerned mainly with elements and non-ferrous alloys, whereas Mode II and III profiles tend to contain some steels and compounds. For the other concepts however the percentage of entities and properties decreased from Mode I to Mode III profiles, activities were approximately constant and abstracts and heterogeneous terms increased from Mode I to Mode III. Thus the types of concepts involved seemed to have a greater correlation with the three performance modes than did the quantitative aspects considered previously.

An outline of the profiles, with an assessment of their performance, can be drawn up and the structure is roughly:

Mode I: A short profile with elements or alloys as the material and properties, activities and entities for the majority of the other concepts.

Mode II: A longer profile, including steels or compounds, with approximately the same proportion of activities, smaller proportions of entities and properties and more abstract and heterogeneous terms.

Mode III: A longer profile, including steels, with approximately the same proportion of activities, but even fewer entities and greater percentages of abstract and

TABLE 32

Analysis of profiles by component concepts

Profile	Element	Alloy	Steel	Compound	Entity	Activity	Abstract	Property	Heterogeneous
Mode I									
12	5	-	-	-	5	5	-	3	1
24	-	1	-	-	1	-	-	-	-
3	2	-	-	-	3	2	2	4	-
21	1	-	-	-	6	4	1	1	1
32	1	1	-	-	4	2	3	3	-
35	3	2	-	-	-	2	2	5	-
56	5	-	-	-	-	1	2	6	-
47	4	4	-	1	1	6	2	5	-
Total	21	8	-	1	20	22	12	27	2
%	18.6	7.1	-	0.9	17.7	19.4	10.6	23.9	1.8
Intermediate									
42	2	1	7	-	1	9	2	3	-
%	8.0	4.0	28.0	-	4.0	36.0	8.0	12.0	-

TABLE 32 (continued)

Profile	Element	Alloy	Steel	Compound	Entity	Activity	Abstract	Property	Heterogeneous
Mode II									
7	-	5	-	-	4	6	6	1	-
10	5	-	-	-	-	3	6	3	-
69	-	1	-	-	-	-	-	2	-
14	1	3	-	-	5	7	3	7	1
66	4	-	-	-	4	4	1	6	1
26	-	-	6	1	1	6	3	3	1
9	7	-	-	1	6	3	10	7	-
Total	17	9	6	2	20	29	29	29	3
%	11.8	6.3	4.2	1.4	13.9	20.1	20.1	20.1	2.1
Mode III									
64	2	2	-	-	1	1	6	5	-
13	6	-	-	-	3	2	5	4	1
57	6	2	-	-	2	4	5	3	2
65	2	2	1	-	6	10	10	2	-
Total	16	6	1	-	12	17	26	14	3
%	16.8	6.3	1.1	-	12.6	17.9	27.4	14.7	3.2

heterogeneous terms.

Thus the relative proportions of entities, properties, abstract and heterogeneous terms apparently are the most significant in determining performance and this is not altogether surprising. Entities and properties are often clearly defined terms which the user, system searcher and original abstractor are likely to agree on the meaning of. Abstract and heterogeneous concepts are less clearly defined and human beings may differ in their semantic interpretation of such concepts, which in itself will lead to misunderstandings and incorrect retrieval.

In passing it is worth noting that in this particular subject area of properties of metals, steels can cause considerable problems. The wide range of steels, with slightly differing compositions and hence properties and uses, can cause many problems in the exact definition of the profile, which in turn can result in large numbers of excluded terms relating to specific compositions or trade names. In indexing it was found essential to include trade names of steels, both for specifying precise compositions and also to facilitate the exclusion of "industrial alloys" for several academic users whose interests were related to new "research" formulations.

3.3.4 Organizational groups

The distribution of users across organizational groups was discussed in Section 2.3, in which the proportions of users

from academic, industrial and government research organizations was given. In order to ascertain whether the profiles from the varying groups demonstrated any marked difference in results, an overall performance measure was considered advantageous. The merits of various measures have been discussed in Section 3.2.6 and both the Q value and the product (M x P) have been used. The values are given in Table 13. Ordering the profiles by these performance measures gives no correlation between organization and performance. In fact both the best and the worst performances, by both measures, were obtained with profiles from industrial organizations. Research institutions, academic and the remaining industrial organization profiles were fairly randomly distributed throughout the order.

Averaging the (M x P) product for the three groups produced the following figures:

Industrial	0.614
Research institution	0.623
Academic	0.573

However the difference between the university and research institution profiles is hardly significant. With Q values the following group averages were obtained:

Industrial	0.9949
Research institution	0.9986
Academic	0.9972

With both measures the research institutions gave the best

performance, although the order of industrial and academic profiles was reversed. Again the differences are hardly significant.

Although there were no significant differences between groups, an impression was formed during the experimental part of the research that the industrial profiles were changing more frequently and that the data base was basically more suited to the academic user.

3.3.5 Profile modification

The modification of the profiles was a continual operation, as necessary, throughout the experiment. As a trial run was conducted with the June issue, some of the profiles were first modified after the June selection, whereas others were not modified until after the July selection. The extent of modification is shown in Table 33, which shows that 26 out of 30 profiles were modified after the June selection and 36 out of 43 profiles were modified after the July issue selection. One profile remained static throughout the project and only five other profiles were not amended after their first selection. The number of profiles modified declined after each monthly selection during the project (with the exception of the November issue), but following the December selections 20 profiles still required some adjustment.

The extent of modification of individual profiles is shown by the frequency distributions given in Table 34. This

TABLE 33

Profile modifications (indicated by 'x')

Profile	Modification after monthly selection						
	7006	7007	7008	7009	7010	7011	7012
1	x	x	-	-	-	-	-
3	x	-	x	-	-	-	-
6	-	-	-	-	-	-	-
7	x	x	x	x	x	x	-
9	x	x	-	-	x	-	-
10	-	x	x	x	x	x	-
12	x	x	x	-	-	-	-
13	x	x	-	-	x	-	-
14	x	x	x	x	x	x	x
21	x	x	-	-	x	x	-
22	x	x	-	x	-	x	x
24	-	-	-	x	x	x	-
26	x	x	x	x	x	x	-
27	x	x	x	x	x	x	x
29	x	x	x	-	-	x	-
30	-	x	x	-	x	x	x
31	x	x	x	x	x	x	-
32	N	-	x	-	x	x	x
34	x	x	x	x	x	x	x
35	x	-	-	x	-	-	-
36	x	-	-	-	x	-	-
37	x	x	x	x	x	x	-
40	x	x	x	-	x	x	x
41	x	x	x	x	-	-	x
42	x	x	x	-	-	x	x
43	N	-	-	-	-	x	-
47	x	x	x	x	x	x	x
51	N	x	-	-	-	-	x
53	x	x	x	x	x	x	x
54	N	x	x	x	x	-	x
56	x	x	x	x	-	-	-
57	N	x	x	-	-	-	x
58	x	x	x	-	x	x	-
59	N	x	-	x	-	x	-
60	x	x	x	x	-	-	-
61	N	x	-	x	-	x	x
62	x	x	x	x	-	x	x
63	N	x	x	-	-	-	x
64	N	x	x	x	x	x	x
65	N	x	x	x	x	x	x
66	N	x	-	x	-	-	x
67	N	x	-	x	-	x	-
69	N	x	x	x	-	-	-
Modified	86.7%	83.7%	65.1%	58.1%	51.1%	60.5%	46.5%

'N' indicates no selection made

TABLE 34

Frequency analysis of profile modifications

Profiles modified after selection for:	No. of profiles	Cumulation
7 issues	4	4
6 issues	9	13 (31%)
5 issues	6	19
4 issues	6	25
3 issues	11	36 (86%)
2 issues	5	41
1 issue	1	42
0 issue	1	43

demonstrates that 86% of the profiles were modified after half the selections and 31% after all, or nearly all, the selections. Several of the profiles were modified intermittently and it did not necessarily follow that the profiles still being adjusted after the December selection had been modified continuously throughout the period.

The quantification of the type of modification is complex. On the whole their nature was becoming less drastic and was gradually being restricted to minor details. Some of the modifications towards the end of the experiment were due to changing interests and it was not possible to differentiate in all cases between changing interests and correction of misinterpretations, etc. It was therefore a disappointing aspect of the research that

one was unable to conclude that the profiles had reached a stabilization point during the project. Only 14 profiles gave the appearance of having more or less stabilized and the majority of these users were from academic organizations. There was no correlation between the extent of profile modification and mode of operation.

The evaluation of a system with profiles that are not initially stabilized is perhaps a rather unfair test and results such as these emphasize the difficulties of the test situation. The experimental part of the project could perhaps have been fruitfully prolonged, but there are no data for the average length of time to stabilize and it is difficult to differentiate between misinterpretations and changing interests. If the selections against the data base were made more frequently than monthly, it might be possible to distinguish between the types of modification. There is scope for further investigation on these aspects, together with methods of obtaining an accurate statement of the users' information needs.

3.3.6 Comparison of two similar profiles

The relationship between the user's statement of interests and the user's actual information needs is an aspect of ISR system testing that is crucial to the success of any system, but very often particularly difficult to investigate. An opportunity presented itself for a very small, but quantitative, investigation of such a relationship. Two of the users, working

in the same organization on closely related topics, gave one initial profile on the claimed basis that their interests were identical. Throughout the experiment it became more and more apparent that their interests, as represented by the documents selected from the total collection as relevant, were related but distinct, and consequently the initial profile was modified to suit their individual requirements more closely.

To give a measure of the similarity between the two profiles a simple consistency measure was used. The measure is the ratio of the number of items selected by both users to the total number of different items selected and can be represented as

$$C = \frac{n(T_1 \cap T_2)}{n(T_1 \cup T_2)}$$

where T_1 and T_2 denote sets selected by the first and second user respectively and n(T) denotes the number of elements in the set. The results for the seven months (both users took part in the trial run with the June issue) are given below:

Issues	Consistency (C)
June	0.382
July	0.500
August	0.273
September	0.250
October	0.400
November	0.222
December	0.304

From these figures it can be seen that the consistency ratio is not very high in any of the months and it is very variable. Also

there is a slight tendency for it to fall off during the period. This does give some indication of the difference between the statement and the actual needs of the user.

3.3.7 Size of user selections

It has been suggested [50] that there is a maximum limit to the number of items that a user will select from a data base and that this limit will be operative irrespective of the number of relevant or partially relevant items in the collection. Such a limit would be imposed by the physical reading capabilities of an individual user and hence may vary from one person to another. From the viewpoint of operating SDI services, such a limit would be important in so far as it should be possible to disseminate to him, say, the ten most relevant items, not just any ten relevant or partially relevant papers from the collection. If it were possible to accurately assign relevance values to items, then it would demand a different approach to the testing of efficiency of SDI systems.

It is of interest to investigate the number of relevant papers selected by users, i.e. the size of C. The range of figures is quite large, varying from 0 to 76, the largest number of items selected from one issue by one user. (It should be noted that an information officer made the latter selection; the largest number selected by a bench scientist was 46.) The average number of items selected from the six issues is given in Table 35, together with the values of N and average generality. The overall average number

of items selected per month was 9.4, although the actual figures varied from month to month depending on the size of the issue, which gave approximately constant values of average generality.

TABLE 35

Average number of items selected from issues

Issue	Average number of items selected	N	Average generality
7007	14.1	723	0.0195
7008	8.0	448	0.0179
7009	6.2	289	0.0215
7010	6.0	298	0.0201
7011	11.4	587	0.0194
7012	10.4	480	0.0217
Total	56.6	2825	0.0200

Average number selected per month = 9.4

This would indicate that the number selected bears some relationship to the size of the collection, at least with values of N of this order of magnitude. It may be that with large document collections, the average value of C does reach a maximum which is independent of N. The result is similar to that obtained by the INSPEC investigation [13] in which the data did not support the suggestion that users tend to mark a smaller proportion of documents relevant when presented with larger numbers of notifications.

For individual users the average number of abstracts selected was also studied and this gave the frequency distribution shown in Table 36. This shows that on average 28% of users selected between 0 and 5 abstracts per month and 47% of users selected between 5.1 and 10 abstracts per month. In other words about three-quarters of the users selected less than 10 abstracts per month.

TABLE 36

Frequency distribution of average number of abstracts selected per month

Average number of abstracts	No. of users	%
0 - 5.0	12	27.9
5.1 - 10.0	20	46.5
10.1 - 15.0	2	4.7
15.1 - 20.0	4	9.3
20.1 - 25.0	2	4.7
25.1 - 30.0	1	2.3
30.1 - 35.0	1	2.3
35.1 - 40.0	-	-
40.1 - 45.0	1	2.3
	43	100.0

The amount of variation is also of interest and an analysis is given in Table 37. In fact 51.2% of users had a selection range of 0 - 10 abstracts and 32.6% of users had a range of 11 - 20. Five users (11.6%) had ranges of 21 - 30 and

their average monthly selections were 34.1, 26.0, 24.1, 17.1 and 8.3. The users with ranges of 31 - 40 and 51 - 60 were both users who selected large numbers of abstracts, on average 22.7 and 41.0 per month respectively. This analysis shows that the users are relatively consistent in the number of abstracts which they select.

TABLE 37

Frequency distribution of range of number of abstracts per month selected

Range	No. of users	%
0 - 10	22	51.2
11 - 20	14	32.6
21 - 30	5	11.6
31 - 40	1	2.3
41 - 50	-	-
51 - 60	1	2.3
	43	100.0

3.3.8 Relevance

In the testing of ISR systems the problem of relevance cannot be ignored, even though discussion of it appears to do little to effectively improve the situation. This subject has probably had more words written about it than any other single aspect of information science and the problem is still there and no realistic methods of handling it have yet been suggested. To attempt to survey the literature of this subject would be a project in its own right and

the resulting gain would probably not justify the time spect, in view of the previous efforts by other people (e.g. Saracevic).

Literature discussing relevance first appeared in about 1964 when Goffman [43] showed by set theory that a relationship between a document and a query is necessary but not sufficient to determine relevance. The two major investigations into the concept have been conducted by the Case Western Reserve University and the Systems Development Corporation (SDC), both of whose studies ended in 1967. Cuadra and Katter at SDC [22] listed 38 variables relating to five aspects of relevance judgements, namely the document, judgement conditions, information requirement statement, the judge and available mode of expression. About half of these variables were examined and it was concluded that all of them can influence the judgement. The Case Western Reserve University study, led by Rees and Schultz, was hardly more conclusive and was closely related to the relevance studies incorporated into the Comparative Systems Laboratory project at the same University. As part of this work, Gifford and Baumanis [42] concluded that user choices were more closely related to co-occurrences of words in the text than had been supposed, Rees and Schultz concluded among other things, that stable meaningful judgements would result only when a group of judges perform the task of evaluating relevance. A somewhat similar approach was adopted by O'Connor]71] in whose study two judges compared 82 documents and 30 questions to find answer-providing documents. There were 32 independent agreements and 48 disagreements, which were resolved by discussion.

There are of course two aspects of relevance studies to consider, the first is the study of relevance per se and the second is the effect of relevance assessments in ISR system evaluation. Considering the latter aspect, Lesk and Salton [63] used a collection of 1200 documents to study the effects on recall and precision of relevance assessments and found that large scale differences did not produce significant variations in average recall and precision. Cleverdon [16] has also investigated the effect of relevance decisions on the Cranfield II results and concluded that the qualitative results would not be changed, i.e. the order of performance of the various languages would be largely unaltered, although the actual percentage results might be.

So the problems of relevance assessment are still with us, nearly ten years after the initial explorations into the area. It seems that in ISR system testing one can either accept the relevance decisions of the users on their face value, incorporating all the human judgement factors that are involved. and hope that as a result of the averaging procedure the effects of erroneous decisions will be smoothed out, or one is overwhelmed by this multitude of variables and concludes that ISR system testing should cease until such time as an alternative testing methodology, not requiring the use of relevance decisions, has been devised. The former is the more positive approach to the situation, although it does raise the question of how large a system is necessary in order that the variations do average out. The third approach involving a team of at least two judges who should reach a joint decision on the relevance or otherwise of the document to the question, has

certain virtues, if an adequate supply of judges is available, but this approach is open to the criticism that it is not representative of a real life situation in which one user requests an output from a service provided perhaps by one operator.

This discussion of relevance is itself pertinent to the magnitude of selection of change by various users during the present project. The methodology employed involved the selection of the relevant subset, C, from the total collection. After comparing the documents in C with those in L (the selected documents), some of the documents were returned to the user for re-examination and comments on selection or non-selection. The comments ranged from "missed it" to "don't know why I selected it", indicating degrees of change in relevance assessment. This procedure has its own dangers in so far as stochastic processes come into effect and the user conception of relevant documents may have altered, in a way similar to that shown by Goffman [44]. However, it is noteworthy that the extent of change in relevance decisions varied from those users who changed a relatively large number of their assessments every month, to those who stuck steadfastly to their original judgement. In fact only two users fell into the latter category for the entire duration of the project and for one of those, there were no relevant papers in the data base.

So the remaining 41 users changed their opinion at some point during the project. The extent of change is shown in Table 38, from which it can be seen that 12 users (28% of the sample) changed their selection in every case, 7 users in five of the six

TABLE 38

Analysis of selection change

No. of months in which a selection change was made	No. of users making the selection change
0	2
1	3
2	3
3	5
4	11
5	7
6	12

months and another 11 users in four of the six months. Thus 70% of the sample (30 users) were being inconsistent in their selections during at least two-thirds of the project.

Another method of assessing the uniformity of users' selections is to use a consistency measure of the type

$$C = \frac{n(T_1 \cap T_2)}{n(T_1 \cup T_2)}$$

where T_1 and T_2 denote the sets of first and second selections made and $n(T)$ is the number of elements in the set. An overall consistency ratio, i.e. the cumulative ratio over the six months of the project, was calculated for each user and the results are tabulated in Table 39. A frequency distribution of the results is shown in Table 40. From these tables it can be seen that about one third of the users were making reasonably consistent

TABLE 39

Overall user consistency ratios

Profile	Ratio	Profile	Ratio	Profile	Ratio
1	0.625	29	0.713	53	0.869
3	0.615	30	0.889	54	0.857
6	-	31	0.672	56	0.810
7	0.702	32	0.522	57	0.821
9	0.931	34	0.588	58	0.500
10	0.500	35	1.000	59	0.690
12	0.845	36	0.585	60	0.719
13	0.925	37	0.632	61	0.393
14	0.645	40	0.843	62	0.333
21	0.581	41	0.795	63	0.640
22	0.735	42	0.681	64	0.690
24	0.898	43	0.545	65	0.844
26	0.648	47	0.761	66	0.873
27	0.817	51	0.731	67	0.833
				69	0.875

TABLE 40

Frequency distribution of user consistency ratios

Range of ratio	No. of users
0.000 - 0.100	-
0.101 - 0.200	-
0.201 - 0.300	-
0.301 - 0.400	2
0.401 - 0.500	2
0.501 - 0.600	5
0.601 - 0.700	10
0.701 - 0.800	7
0.801 - 0.900	13
0.901 - 1.000	3

assessments. However, 17 users had ratios in the range 0.601 - 0.800, indicating that about one-third of their assessments were subject to change and four users had ratios less than 0.500, indicating low consistency.

The consistency of the users varies from month to month, although in general the degree of change increased towards the end of the experiment, as shown by the monthly averages of the ratios given in Table 41. This could be due to a great many factors including all those which affect relevance judgements. In addition the increasing time lag between the users making their original assessment and their revised assessment as the experiment progressed

TABLE 41

Average consistency ratios

Month	Average
July	0.785
August	0.744
September	0.704
October	0.671
November	0.655
December	0.681

made it more likely that the larger degree of change would occur with the later issues. This delay varied widely from user to user as some returned their comments on the selections quickly and thus could be progressed rapidly through the procedures, whereas others had to be continually followed up for their comments. Some idea of this variation can be obtained from the facts that the first selections against the December 1970 issue were made on the 15th June 1971 and the last was made on the 18th April 1972. In fact the last set of comments were not received and processed until the 11th May 1972.

The lack of consistency in relevance assessments made by the users in this project emphasizes both the difficulty of obtaining significant relevance judgements and the effects of stochastic processes. In projects such as this where individual values of the primary parameters tend to be small in many instances,

a doubtful relevance decision can have a disproportionately large effect on the calculated ratios.

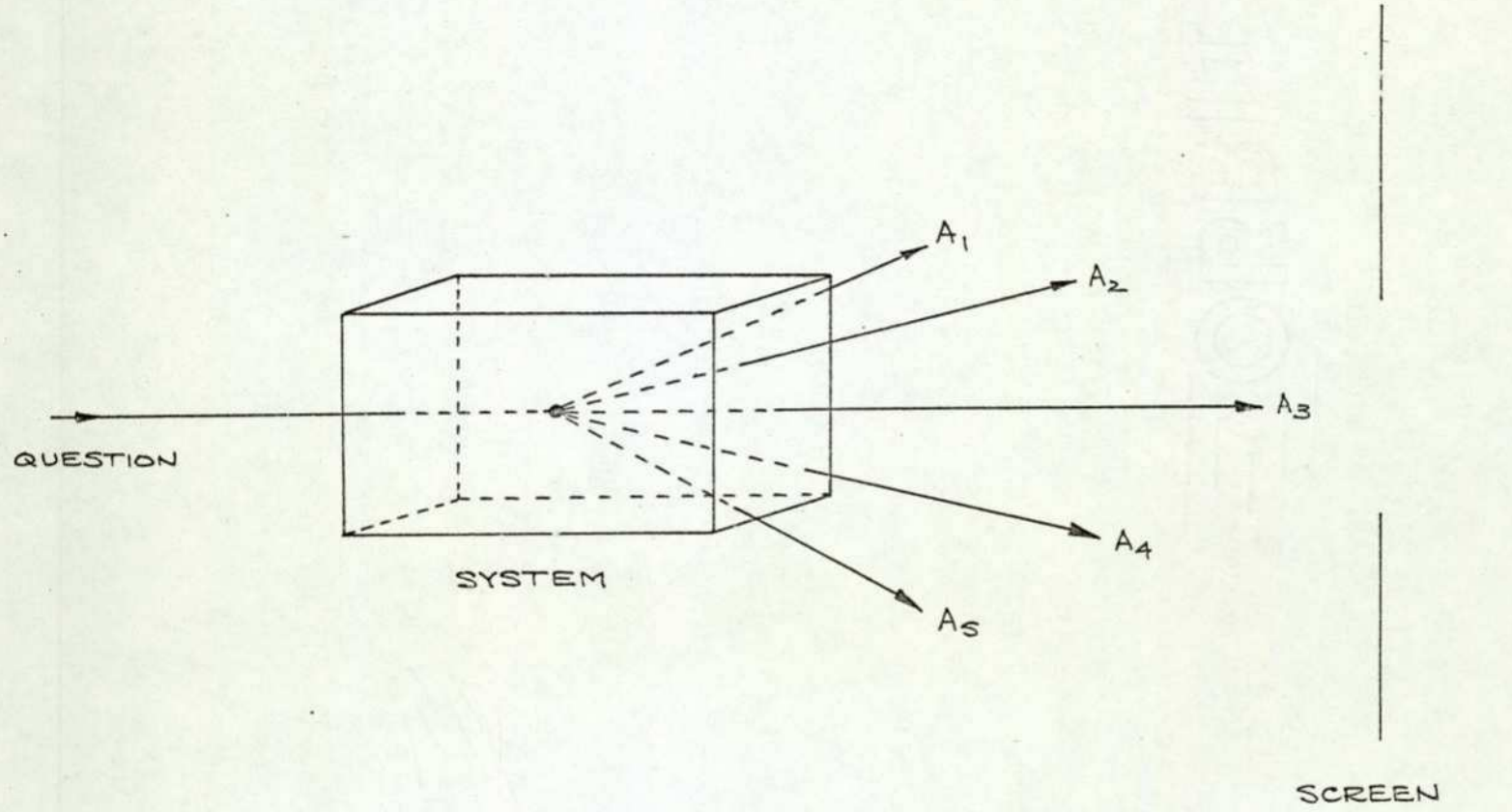
4. GENERAL DISCUSSION

4.1 Methodology

In discussing methodology in Section 2.1 the comment was made that an information storage and retrieval system as used for testing incorporates such a vast number of uncontrolled and uncontrollable variables that they probably mask any true performance factors associated with the system. The problem of circumventing this situation remains.

An ISR system could be depicted by a model (shown in Fig. 56) in which a question is input into the system. The interaction between the question and the system results in a number of answers, A_1, A_2, A_3, A_4, A_5, etc., which are output in various directions depending on the nature of the interaction. The user then acts as a screen or filter on this output, allowing some answers through as relevant and stopping others as non-relevant. The screen can have an aperture of varying sizes and positionings in relation to the system. These parameters represent the varying facets of the user. In fact some users act as if they were sieves, letting through a random selection of answers!

Such a model, although a rather crude representation of a system, serves to illustrate some aspects of testing methodology. The variable aperture size and positioning of the screen emphasizes that the output or answers should be evaluated or measured in some way before they reach the screen or, alternatively, that they should be evaluated by means of standard, or a series of



MODEL OF A GENERAL INFORMATION STORAGE AND RETRIEVAL SYSTEM

FIGURE 56

standard, screens. Thus a departure from the user or judge situation in which each user poses a question and assesses the response is necessary. In an SDI situation the variables are more extreme than in a retrospective search because the user is bringing them into play with each set of responses and, because the responses are made at different times, they may lack consistency for any one user. In terms of the model the screen may be constituted differently on occasions. A panel of judges which comes to a joint evaluation of the output, as recommended by O'Connor [71], has certain advantages in that they eliminate some of the variables in coming to their joint decision. A panel of judges could be likened to a standard screen. However, this technique is one step away from the real life situation. The fact that judges working under different conditions will rank output in more or less the same order [16] is helpful in that it implies that there is some basis for relevance decisions, even if the fact that they may give them different ratings implies that a quantitative result is not realistic.

If the output is not assessed by standard screens, it should be evaluated by some other means before reaching the screen. This aspect is more difficult as the majority of testing situations are measured by recall and precision or other measures which are related to these. Many people's thinking is restricted by the dominance of this type of measure and the emergence of a spark of creativity, or the application of de Bono's "lateral thinking", will be necessary to overcome this barrier, if indeed it can be surmounted.

The model in Fig. 57 symbolises the ISR system as a "black box" concept. But the human aspect of the system must be considered. Any system, apart from those manipulating original text by computer, must involve some measure of human intervention and thus be subject to error. But this is part of the system itself and thus efforts will have to be made to isolate the factors and control them.

A third variation to consider is a statistical approach. If it is assumed that it is not possible to control all the human aspects, especially those incorporated within the system itself, then perhaps tests should be designed to include sufficient individual tests to make a truly statistical approach realistic. The main disadvantages militating against such an approach include the difficulty of obtaining enough volunteer users to provide reliable statistical results and, given adequate users, the difficulty of providing a large enough team of investigators to evaluate the system.

Perhaps the real answer is an entirely different approach to testing. Although even then it is difficult to visualize a new one which would satisfy both scientific method and the proponents of the "real life situation". At present the tester is in a cleft stick. If the control of some of the variables is attempted, by carrying out the experiments in a laboratory type environment, the criticism is made that the system does not adequately reflect the real life situation. If, on the other hand, a real life situation is chosen, the disadvantages are

that it may not be practicable to measure all the parameters necessary for a quantitative evaluation, and also the myriads of uncontrollable variables may lead to both dispair and the criticism that the experiments are unscientific and unreproducible. It would seem likely that information scientists will have to agree that, in the first instance, a scientific, laboratory environment may be necessary for progress in deciphering the operation of systems or subsystems, with a later extension of the methodology into the field, when there is some sounder basis on which to conduct tests.

Returning to the model, the present situation is one in which questions of a more or less random nature are bombarded at the system. A response is achieved which is probably the amalgamation of the various question - system interactions, which is vastly magnified by the questions, and hence the interactions, all being different. Thus it seems optimistic to anticipate that the output will show any discernible pattern. Perhaps the first stage should be to classify questions and ascertain whether different responses are achieved. The second stage may be to control more completely the data base and to include evaluated items. It has been suggested by Farradane [35] that seeding known evaluated items into the system and ascertaining whether or not these are retrieved in response to specific questions may be the only way of overcoming the user relevance assessment. Of course there are difficulties to be overcome in building up a suitable data base and questions in adequate numbers. The study of the individual sections of the total system would seem to offer

a more fruitful approach than the somewhat random techniques at present adopted.

If the study of a system as depicted by the model permits the understanding of the type of responses obtained under certain conditions, then the way is opened for the better design of systems. A parallel study of the screens (users) would perhaps permit the question - system interaction to produce answers directed at the apperture in any one screen. For example, if it could be established that one user represented a screen with a certain type of aperture in a particular position through which say A_2 responses pass, then perhaps the system could be operated in such a way as to produce A_2 responses only. However, this situation seems to be rather a long way in the future.

The discussion of models of ISR systems is interesting, but perhaps the root causes of the testing problems are not as far away as investigators wouldlike to think. Snyder et al.'s report [92] on methodology reviews and analyzes 49 studies of document retrieval systems. As a result of this analysis, eight methodological problems were identified, namely (1) confounding of independent variables, (2) using only retrieval effectiveness criteria to examine effects of independent variables, (3) using gross measures to determine relevance, (4) using insufficient control of possibly significant sources of variance, (5) using inadequate sampling procedures, (6) using inadequate statistical procedures and techniques, (7) failing to acknowledge or discuss other relevant research results, and (8) failure to provide enough detail when

reporting methodology of a study.

One project which presumably would not be subject to such criticism is the work of the Comparative Systems Laboratory at Case Western Reserve University. Saracevic [87] has given a condensed report of the results of this project, from which the main conclusion was that the human factor, the variations introduced by human decision making, is the overwhelming variable. Richmond [79] in her review of the Final Report, comments that Saracevic and his team "attempted to apply scientific method with insight and sophistication to obtain a closer approximation to truth. They have been exhaustive in the attempt to define all the variables, which was not entirely successful because more variables kept turning up." The overall conclusion of the work was that quantitative evaluation was possible, but only after much basic research and much more study.

Again, perhaps the problem is one of outlook. By the nature and history of information science, the majority of system testers in the past have been scientists, particularly physical and biological scientists. Thus the techniques brought to bear upon the problem have tended to be similar to those applied to scientific problems. Some six years ago, Rees and Schultz [75] proposed that the behavioural sciences, specifically psychology, should be applied to information science, in the area of methodology. The Comparative Systems Laboratory project also concluded that the problems involved are essentially those of the behavioural sciences, and that evaluation studies could be expected to follow the progress

in the fields of psychology, sociology, cultural anthropology, etc. Perhaps therefore it would be fruitful to investigate these areas or to recruit some behavioural scientists into the research teams.

4.2 User needs

The behavioural sciences are of course being utilized in information science already. The relational indexing system itself is based on the psychology of thinking. Additionally it is thought that the psychology of the user could prove helpful in such aspects as the selection of items from abstracts journals.

Indeed perhaps the information sciences themselves should be classified as behavioural sciences because the interface between the system and the user, the flow and communication of information, are so dependent upon user behaviour. The approach is however being incorporated, as exemplified by the work on relevance. One of the factors that affects relevance decisions, as shown by Saracevic and his colleagues [88], is the stage of research. In the early stages a user will accept many papers, either because he wants to obtain an overview of the problem or because the problem is not fully defined in his own mind. As the research progresses he becomes much more discriminating in his choice of items. The situation is similar to the "hierarchy of needs" developed by Maslow [65], in which it is proposed that people have certain basic (physiological) needs (for food, drink, shelter, etc.) and only when these needs are satisfied will a person become concerned about satisfying his needs for safety, social acceptance, self-

esteem and self-realization, working up the hierarchy. Thus the urge to satisfy these needs will motivate a person's behaviour.

Maslow's motivational theory and also Herzberg's Hygiene - Motivation Theory [53] have greater implications for information science. If a service is to be provided to users by means of an ISR system, then these motivational factors should be incorporated into the system in order that it can more adequately meet their needs at any stage.

A move was made in this direction by Harmon [51]. He has re-interpreted the findings of the Cuadra - Katter and Rees - Schultz experiments on relevance in terms of a user information need model, which he likens to a mental jigsaw. A user acquires information which he orders; the gaps are filled and adjustments made to the ordering. Gradually the user becomes more selective because his needs are better circumscribed and the problem is ultimately one of gap-filling rather than ordering.

A further manifestation of user needs is the quantity of literature read by users. A little research has been carried out on the maximum number of papers that a user would wish to receive and read, on the hypothesis that there is a fairly small number that represents the number of papers per week, say, that one person could handle. Cleverdon and Harding [50] considered this aspect for precision engineers and surmised that 20 papers per fortnight would be an optimum number. A survey by Parker [72] at Southampton University indicates slightly lower figures in that

the majority of users read 1 - 2 articles per week in depth, could handle a maximum of 3 - 10 articles per week in depth, and scan 3 - 5 articles per week. Thus for an ISR system to be truly effective, it should select the three (or some other number specified as an individual threshold) most pertinent articles per week.

4.3 Users' comments

Some of the volunteer "users" of the relational indexing SDI system made comments during the course of the project. Although the majority of these say nothing radically new or different, some of them are reproduced here to illustrate various points.

Some of the human factors involved are indicated by the following comments:

User A "Really, to summarise, unless a profile was so detailed that it covered several hundred pages, then I don't think one would ever get an exact one to one relation between selection from the profile and selection by my hand. I have no doubt that a re-examination of the abstracts journal in one year's time would lead me to select a somewhat different set of abstracts.

There is also the human operator error in scanning through a wad of abstracts - one can miss a paper of interest just through fatigue, or untypical titling."

User B "I hope these brief notes will help you to readjust my profile. It seems to me to be necessary to continually adjust my understanding of the words 'interesting' and 'relevant', in a sense anything is interesting which causes mild amusement, or in extreme, 'why didn't I blooming well think of that' type of reactions. It continually crosses my mind that you are attempting an improper classification in the sense that 'relevance' and 'interest' are personal and you are going to judge everybody by their own eyes. Still, you can but consider any individual's use of these abstracts as the grand criterion. In this connection, it may be that your periodic letters provoke a more consistent response in me, and probably others, than under unperturbed circumstances. I can't claim to be the most regular person in consulting abstracts and other sources of information."

User C "I see a number of difficulties.

1. Profile could be improved. I know this is basically my fault, since I didn't reply earlier in order to allow you to modify my profile. This clearly could be different and better with a better organised and more co-operative client. But there is still a basic problem of getting the researchers to specify their areas of interest. I think I must be very bad to doing this, but the fact is that there are not sharp boundaries for various reasons:

   (a) one's interests shift, change, develop with time. I guess this could be coped with by updating the profile frequently.

   (b) a hierarchy of interest. You have provided

us with two levels of interest, and I think it might be useful to say a little about how I determine the level. First, there are things on which I am working directly. I clearly have to have all published information, in any language. Secondly, there are things, properties, ideas, which are strongly linked to what I am doing but don't reproduce it, e.g. structure of solid or liquid Hg. Thirdly things of some but not strong interest, e.g. all those papers on surface tension and viscosity of liquid alloys. As I said, there are just too many of these papers for me to make much of an effort to know about a paper that I have to have translated before I can even read it. On the other hand, I would be strongly interested in a first class critical review article in this field, which would do all the hard work but enable me to get into the field if I wanted or needed to do so for some reason. There are a number of areas like that - strongly interested in reviews or books but only weakly interested in individual papers.

2. Finally, there are really way out things that suddenly, when one reads the abstract, just click as one sees the relevance and interest in a way which I don't think any profile could deal with. I don't see any way round that one."

User D "Many (about 50%) of the abstracts from this section [3.4] could be included in the possibly relevant category."

Language of the article often provides problems:

User E "The choice of abstract depends on its availability, e.g. if an article is in Russian or Japanese, the abstract has to be really worthwhile before a translation is commissioned."

User F "Also a lot of abstracts are from foreign language journals. It is sometimes difficult to decide from the abstract whether or not it is worth getting a translation."

The nature of the user's work and also his location within an industrial or academic environment influence his reactions:

User G "Their [the users] interests tend to be for practical rather than theoretical papers."

User H "He is not interested in any molten salt work and considers that many of the abstracts of a highly specific nature are not of interest to him."

User I "They are working on a variety of projects that come their way - both long term projects and short term projects (which may be from other companies and may only last a week). Therefore they have to be interested in everything without being able to foretell what their interests will be."

Finally, comments of a general nature:

User J "Having gone through the abstracts again I do not feel that the discrepancies are as bad as they at first appeared."

User K "I would say that if I were paying pro rata for abstracts received, then I would not object to any in your list of those which I did not select."

5. CONCLUSIONS

Unfortunately the lack of generally accepted scientific laws, together with compatibility of meaningful methods of presenting results, render comparison between the results obtained in this research and those of other researchers using different systems and techniques extremely difficult. Similarly conclusive evidence that relational indexing is either superior or not to the more practised and established methods in the SDI situation must await further research and study. The results, in terms of recall, precision and fallout, were encouraging and mean system performance, as represented by the XY3 sets, varied between 0.69 and 0.82 for recall, 0.72 and 0.87 for precision and 0.004 and 0.003 for fallout. One of the more encouraging aspects was the attainment of recall and precision ratios of approximately equal magnitude.

From a methodological viewpoint, it was disappointing, but perhaps understandable, that a clear indication of profile stabilization was not achieved. The problem of stabilization compared with interest changes in a "real life" situation is particularly pertinent to SDI system testing. Overall performance measures were inconclusive in this research. The relationships between parameters was investigated. Recall - fallout, recall - generality and precision - generality all appeared to have some sort of hyperbolic relationship and of these, surprisingly, the recall - generality relation was the most convincing.

The distributions of the performance parameters indicated three or four modes of operation. Attempts to correlate

these modes with characteristics of the system were not completely successful, the closest correlation appearing with the qualitative profile content. Other indications were observed that the actual concepts occurring in the profile play a crucial role in the retrieval process.

An error analysis was conducted, 27 error categories being used. User errors were 48% of the total and acceptable system errors a further 15%. Of the non-acceptable system errors, technical implication (6.5%), unspecified (6.2%), generic browsing/omission (4.3%) and conceptual browsing/omission (4.3%) occurred most frequently.

Although first applied to sugar technology, the applicability of relational indexing to other subject areas has been demonstrated by a number of people over a period of years. In this case no difficulty was experienced in using the technique to index abstracts in the field of properties of metals. Certain new situations occurred but the relations were adequate to handle them. In an operational situation it is considered that a classification, or at least a structured list of concepts, is a necessary adjunct to the system to facilitate browsing. Additionally certain conventions within a subject area are required to assist the expression of meaning, for example, the name of an industrial alloy should be specified to indicate that the alloy is industrial and to specify most precisely its compositions.

Additional uses of the relations themselves were encountered. Thus the self-activity relation (/*) was used for "through" and for phenomenological properties. A distinction between past and potential processes may be convenient in which case the dimensional relation (/+) is suggested for past processes and the concurrence relation (/θ) for potential processes. The number of relations associated with one concept does not appear to be limited, nor does their distribution around the concept. The representation of information in the form of concepts and relations can become complex and on a few occasions the need for a three-dimensional format was indicated, as was depicted in Section 3.1.3.

Browsing techniques are necessary adjuncts to information retrieval by relational indexing in order that recall should be maximised. The techniques employed were generic browsing, conceptual browsing, related browsing and condensations (logical jumps). Further techniques such as expansions and inference steps may also be desirable, but their development was inadequate for use in this research. Generic browsing was the most frequently used strategy. In the majority of cases all the techniques led to improved recall with very little deterioration in precision. The usefulness of generic browsing in this research is attributed to the difference in specificity of the concepts used to index the abstracts and the profiles, combined with a more unorthodox approach to applying the other techniques, especially conceptual browsing.

A number of areas of further research were indicated during the project, of which the further analysis and development of conceptual browsing techniques was one. Condensations, related browsing and the significance of concepts not participating in browsing operations warrant further investigation. User needs, in particular obtaining an accurate profile of interests, the correlation with various profile structures and the length of time for profile stabilization, could be studied in greater detail, together with more general research into an individual's requirement for information and the factors that motivate his selection habits. Inter-indexer consistency and methods of ranking output with relational indexing could also be elaborated.

The problems of semantics are not completely overcome with relational indexing. The final stage in an SDI system is the transfer of information to the user and thus communication with him. Drucker [24] points out that communication involves perception, expectation and involvement and is the successful perception of a message or information by the person to whom it was addressed. The communicator is the person who receives, not the person who gives, the information, and the communication will be received only if the recipient is capable of perceiving it. Cherry [12] gives some of the difficulties of achieving a successful communication and Farradane [34] re-emphasizes some of these points by saying "we achieve understanding by the addition of our existing stock of knowledge in interpreting the communication". The difficulties of achieving this final step in the SDI communication chain, the problems associated with

the other factors that affect the interpretation of meaning, have been recognised by Rowlands and Garvey, Lin and Tomita. Rowlands [82] concluded from his evaluation of the Unilever SDI system that the

> "mark of a successful SDI system should surely be its flexibility in catering for all types of user requirements, preferably the user manipulating his profile in such a way as to achieve the required output. The user must therefore be educated in profile construction and to some extent in system operation."

Garvey et al. [41], studying communication among scientists, concluded

> "One third of the authors in our studies did not pursue further work in the same subject matter areas as their articles, but subsequently moved into new areas. We interpret these data to suggest (a) that, when one examines the collective scientific endeavour in a discipline through the activities of individual scientists, one finds in any progressive science a significant amount of discontinuity of individual research work, and (b) that this discontinuity is an essential attribute of the collective progress of science. Further, we detect that the communication problems associated with this discontinuity are time-consuming for the individual scientists whose information needs both change as they progress through the different stages of research, and take on new meanings when they shift into new subject-matter areas. This observation suggests that systems of selective dissemination of information must be remarkably flexible to be genuinely

effective in satisfying ongoing information needs."

These quotations should not be interpreted as a sign for SDI system organisers to retire from the fray, but rather as an indication that the problems associated with the construction and maintenance of accurate representations of the user's information requirement and also the final communication step are crucial. They summarise some of the difficulties that must be acknowledged and overcome if SDI systems are to become truly efficient in an operational situation, and they are amply supported by the experience of this research. The provision of SDI services cannot be based solely on decisions relating to the index language used and the hardware employed for the retrieval process, nor can the evaluation of such services depend only on recall, precision and similar measures of performance. Similarly cost and coverage, although obviously important aspects of a commercial service, will not assess whether or not the user's information needs are met. Because of the human involvement and the consequent human factors, the inclusion of a semantic element, such as that provided by relational indexing, is essential if an SDI system is to attempt to approach maximum recall and precision with minimum fallout. The potential of relational indexing, together with fundamental human behavioural aspects and particularly more appropriate evaluation techniques, should continue to be further explored and developed.

ACKNOWLEDGEMENTS

This thesis owes much to the assistance and encouragement given by the author's supervisor, Mr J. Farradane, (Director of the Centre for Information Science, The City University) throughout the duration of the project. The author expresses her deep gratitude and appreciation, particularly for the stimulating discussions and debates that took place between them, despite the very heavy programme of the Director's commitments in a multitude of other areas.

The excellent secretarial work of [redacted] and [redacted] was also very much appreciated and is gratefully acknowledged.

The author is also appreciative of the kindness and patience extended to her by her lecturing colleagues and the students of the Centre for Information Science (The City University) during what proved to be a most demanding three years.

Finally she also wishes to record her most sincere thanks to all those who volunteered to take part in the project as "users", and most particularly to the 43 people named below who participated in the project for its total duration. Without their help the research could not have been undertaken or completed. It is impossible to record in words the total value of their contribution and the author is all the more grateful for their help, recognising that it made an additional demand on their scarce time. In alphabetical order they are as follows:

[redacted], University of Wales Institute of Science and Technology
[redacted] Lanchester Polytechnic
[redacted], Imperial Metal Industries Ltd
[redacted], University of Exeter
[redacted], Atomic Weapons Research Establishment
[redacted] and [redacted], Tin Research Institute

[redacted], British Aircraft Corporation
[redacted] British Aircraft Corporation
[redacted] Imperial Metal Industries Ltd
[redacted], Brunel University
[redacted], Standard Telecommunications Laboratories Ltd
[redacted], Tube Investments Ltd
[redacted], Imperial College of Science and Technology
[redacted], Atomic Weapons Research Establishment
[redacted], Lanchester Polytechnic
[redacted] Johnson, Matthey and Co. Ltd
[redacted], Royal Military College of Science
[redacted], Royal Aircraft Establishment
[redacted], University of Manchester
[redacted], University of Cambridge
[redacted], Standard Telecommunications Laboratories Ltd
[redacted], The Post Office
[redacted] Joseph Lucas Ltd
[redacted], University College of North Wales
[redacted], Royal Military College of Science
[redacted], Fulmer Research Institute
[redacted], University of East Anglia
[redacted], The City University
[redacted], Warwick University
[redacted], Royal Military College of Science
[redacted], Tube Investments Ltd
[redacted], United Kingdom Atomic Energy Authority
[redacted]. Thomas Bolton and Sons Ltd
[redacted], Joseph Lucas Ltd
[redacted], University of Sheffield
[redacted], Building Research Station
[redacted], Lanchester Polytechnic
[redacted], The City University
[redacted], University of Hull
[redacted], Lanchester Polytechnic
[redacted], Atomic Weapons Research Station
[redacted] The City University
[redacted], University of East Anglia

REFERENCES

1. AITCHISON, T.M. and TRACY, J.M. INSPEC Comparative evaluation of index languages. Part I: Design. London, Institution of Electrical Engineers, July 1969.

2. AITCHISON, T.M., HALL, A.M., LAVELLE, K.H. and TRACY, J.M. INSPEC Comparative evaluation of index languages. Part II: Results. London, Institution of Electrical Engineers, March 1970.

3. ANTHONY, L.J., CHENEY, A.G. and WHELAN, E.K. Some experiments in the selective dissemination of information in the field of plasma physics. Inf. Stor. Retr., 1968, 4, 187-200.

4. AUSTIN, D. Theory of integrative levels reconsidered as the basis of a general classification. In: Classification Research Group, Classification and Information Control, London, The Library Association (Library Association Research Publications No. 1), 1969, 81-95.

5. AUSTIN, D. Development of a new general classification: a progress report. Inf. Scientist, 1969, 3, 95-115.

6. BARKER, F.H., KENT, A.K. and VEAL, D.C. Report on the evaluation of an experimental computer-based current awareness service for chemists. Draft report, February 1969.

7. BARKER, F.H., KENT, A.K. and VEAL, D.C. Report on the evaluation of an experimental computer-based current awareness service for chemists. London, The Chemical Society, Research Report No. 1, August 1970.

8. BARKER, F.H., VEAL, D.C. and WYATT, B.K. Towards automatic profile construction. J. Doc., 1972, 28, 44-55.

9. BERENBLUT, I.I. (The City University) Personal communication.

10. BROOKES, B.C. The measures of information retrieval effectiveness proposed by Swets. J. Doc., 1968, 24, 41-54.

11. BUREK, C.V. Inference steps in thinking by relational indexing as an aid to classification. M.Sc. thesis, London, The City University, 1971.

12. CHERRY, C. On human communication: a review, a survey, and a criticism, 2nd edn, London, The M.I.T. Press, 1966.

13. CLAGUE, P. INSPEC SDI investigation 1967 - 1969. London, Institution of Electrical Engineers, OSTI Report 5102, 1971, 5 vols.

14. CLEVERDON, C.W. Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems. Cranfield, 1962.

15. CLEVERDON, C.W. Evaluation tests of information retrieval systems. J. Doc., 1970, 26, 55-67.

16. CLEVERDON, C.W. The effect of variations in relevance assessments in comparative experimental tests of index languages. Cranfield Institute of Technology, October 1970. (Cranfield Library Report No. 3)

17. CLEVERDON, C.W. On the inverse relationship of recall and precision. J. Doc., 1972, 28, 195-201.

18. CLEVERDON, C.W. and KEEN, M. Factors determining the performance of indexing systems. Vol. 2, Test results. Cranfield, College of Aeronautics, 1966.

19. CLOUGH, C.R. and BRAMWELL, K.M. A single computer based system for both current awareness and retrospective search: operating experience with ASSASSIN. J. Doc., 1971, 27, 243-53.

20. COOPER, W.S. Is interindexer consistency a hobgoblin? Amer. Doc., 1969, 20, 268-78.

21. COOPER, W.S. On deriving design equations for information retrieval systems. J. Amer. Soc. Inf. Soc., 1970, 21, 385-95.

22. CUADRA, C.A. and KATTER, R.V. The relevance of relevance assessments. Proc. Amer. Doc. Inst., 1967, 4, 95-9.

23. DATTA, S. Concept organization for information retrieval. Ph.D. thesis, London, The City University, 1969.

24. DRUCKER, P. What communication means. Management Today, 1970, March, 91-3, 150-1.

25. FAIRTHORNE, R.A. Basic parameters of retrieval tests. Proc. Amer. Doc. Inst., 1964, 1, 343-5.

26. FARRADANE, J. A scientific theory of classification and indexing and its practical applications. J. Doc., 1950, 6, 83-99.

27. FARRADANE, J. A scientific theory of classification and indexing: further considerations. J. Doc., 1952, 8, 73-92.

28. FARRADANE, J. The psychology of classification. J. Doc., 1955, 11, 187-201.

29. FARRADANE, J. Relational indexing in the light of recent experimental work in psychology. Inf. Stor. Retr., 1963, 1, 3-11.

30. FARRADANE, J. Information for design. In: The design method, edited by S.A. Gregory, London, Butterworths, 1966, 97-103.

31. FARRADANE, J. Concept organization for information retrieval. Inf. Stor. Retr., 1967, 3, 297-314.

32. FARRADANE, J. Research on relational indexing. Final (condensed) report to OSTI. London, The City University, 1968. (Unpublished)

33. FARRADANE, J. Analysis and organization of knowledge for retrieval. Aslib Proc., 1970, 22, 607-16.

34. FARRADANE, J. The necessity of semantic analysis for information retrieval. Paper presented at the Aslib Co-ordinate Indexing Group Conference: Informatics I, Durham, 11-13 April 1973.

35. FARRADANE, J. Personal communication.

36. FARRADANE, J., DATTA, S. and POULTON, R.K. Report on research on information retrieval by relational indexing. Part I: Methodology. London, The City University, 1966.

37. FARRADANE, J., DATTA, S. and HILLS, J. Unpublished work.

38. FARRADANE, J., RUSSELL, J.M. and YATES-MERCER, P.A. Problems in information retrieval: logical jumps in the expression of information. Inf. Stor. Retr., 1973, 9, 65-77.

39. FARRADANE, J. and YATES-MERCER, P.A. Retrieval characteristics of Metals Abstracts Index. Unpublished.

40. GARDIN, J.C. SYNTOL, New Brunswick, N.J., Rutgers State University, 1965. (Rutgers Series on Systems for the Intellectual Organization of Information, ed. S. Artandi, Vol. II).

41. GARVEY, W.D., LIN, N. and TOMITA, K. Research studies in scientific communication. IV. The continuity of dissemination of information by "productive scientists". Inf. Stor. Retr., 1972, 8, 265-76.

42. GIFFORD, C. and BAUMANIS, G.J. On understanding user choices: Textual correlates of relevance judgements. Amer. Doc., 1969, 20, 21-6.

43. GOFFMAN, W.A. On relevance as a measure. Inf. Stor. Retr., 1964, 2, 201-3.

44. GOFFMAN, W.A. An indirect method of information retrieval. Inf. Stor. Retr., 1968, 4, 361-73.

45. GOFFMAN, W.A. Personal communication.

46. GOFFMAN, W.A. and NEWILL, V.A. A methodology for test and evaluation of information retrieval systems. Inf. Stor. Retr., 1966, 3, 19-25.

47. GUILFORD, J.P. Three faces of intellect. Amer. Psychologist, 1959, 14, 469-79.

48. GUILFORD, J.P. Basic conceptual problems in the psychology of thinking. Ann. N.Y. Acad. Sci., 1960, 91, 6-21.

49. GUILFORD, J.P. Intelligence, creativity and their educational implications, San Diego, California, Robert R. Knapp, 1968.

50. HARDING, P. Personal communication.

51. HARMON, G. Information need transformation during inquiry: a re-interpretation of user relevance. In: Proceedings 33rd Annual Meeting of the American Society for Information Science, Philadelphia, October 11-15, 1970; Vol. 7: The information conscious society, edited by J.B. North, Washington DC, the Society, 1970, 41-3.

52. HEAPS, H.S. and KO, W.C.C. Automatic adaptive processing of questions in document retrieval. In: Proceedings 33rd Annual Meeting of the American Society for Information Science, Philadelphia, October 11-15, 1970; Vol. 7: The information conscious society, edited by J.B. North, Washington DC, the Society, 1970, 319-21.

53. HERZBERG, F. Work and the nature of man. London, Staples Press, 1968.

54. HOUSMAN, E.M. Survey of current systems for selective dissemination of information (SDI). Washington DC, American Society for Information Science, Special Interest Group on Selective Dissemination of Information (SIG/SDI), June 1969. (AD-692 792)

55. JOHNSTON, J.O.D. Expansion of ideas analysed by relational indexing. M.Sc. thesis. London, The City University, 1971.

56. KEEN, E.M. The Aberystwyth index languages test. J. Doc., 1973, 29, 1-35.

57. KEEN, E.M. and DIGGER, J.A. Report of an information science index languages test. Aberystwyth, College of Librarianship Wales, 1972.

58. KERR-WALLER, R.D. Automated Information Dissemination System (AIDS): technical evaluation. Inform. Sci., 1967, 1, 51-72.

59. KONIGOVA, M. Mathematical and statistical methods of noise evaluation in a retrieval system. Inf. Stor. Retr., 1971, 6, 437-44.

60. LANCASTER, F.W. Information retrieval systems: characteristics, testing, and evaluation. New York, John Wiley and Sons, Inc., 1968, 168-9.

61. LANCASTER, F.W. MEDLARS: report on the evaluation of its operating efficiency. Amer. Doc., 1969, 20, 119-42.

62. LANCASTER, F.W. Aftermath of an evaluation. J. Doc., 1971, 27, 1-10.

63. LESK, M.E. and SALTON, G. Relevance assessments and retrieval system evaluation. Inf. Stor. Retr., 1969, 4, 343-59.

64. MACKENZIE, C.A. Relational indexing in the German language. M.Sc. thesis. London, The City University, 1972.

65. MASLOW, A.H. Motivation and personality. New York, Harper and Row, 1954.

66. McCASH, W.H. and CARMICHAEL, J.J. UDC user profiles as developed for a computer-based SDI service in the iron and steel industry. J. Doc., 1970, 26, 295-312.

67. MEEKER, M.N. The structure of intellect: its interpretation and uses. Columbus, Ohio, Merrill, 1969.

68. MEETHAM. A.R. Communication theory and the evaluation of information retrieval systems. Inf. Stor. Retr., 1969, 5, 129-34.

69. MILLER, W.L. Extension of users' literature awareness as a measure of retrieval. J. Doc., 1971, 27, 125-135.

70. NELSON, P.J. User profiling for normal text retrieval. Proc. Amer. Doc. Inst., 1967, 4, 288-95.

71. O'CONNOR, J. Some independent agreements and resolved disagreements about answer-providing documents. Amer. Doc., 1969, 20, 311-19.

72. PARKER, C.C. The use of external current awareness services at Southampton University. Aslib Proc., 1973, 25, 4-17.

73. PERREAULT, J.M. Categories and relators: a new schema. Rev. Int. Doc., 1965, 32, 136-44.

74. POLLOCK, S.M. Measures for the comparison of information retrieval systems. Amer. Doc., 1968, 19, 387-97.

75. REES, A.M. The evaluation of retrieval systems. Comparative Systems Laboratory Technical Report No. 5. Cleveland, Ohio, Western Reserve University, 1965.

76. REES, A. and SCHULTZ, D.G. A field experimental approach to the study of relevance assessments in relation to document searching. Cleveland, Ohio, Case Western Reserve University, 1967. (PB-176 079, PB-176 080)

77. REES, A.M. and SCHULTZ, D.G. Psychology and information retrieval. In: Information retrieval: a critical review, edited by G. Schecter, London, Academic Press, 1967, 143-50.

78. REIMER, A.B. Use of browsing techniques with relational indexing. M.Sc. thesis. London, The City University, 1971.

79. RICHMOND, P.A. The Final Report of the Comparative Systems Laboratory: a review. J. Amer. Soc. Inf. Sci., 1970, 21, 160-2.

80. ROBERTSON, S.E. The parametric description of retrieval tests. Part I: The basic parameters. J. Doc., 1969, 25, 1-27.

81. ROMERIO, G.F. and CAVARA, L. Assessment studies of documentation systems. Inf. Stor. Retr., 1968, 4, 309-25.

82. ROWLANDS, D.G. The Unilever Research S D I system. Inf. Stor. Retr., 1970, 6, 53-71.

83. ROWLANDS, V.R. Reliability coefficients for retrieval tests. M.Sc. thesis. London, The City University, 1971.

84. RUSSELL, J.M. Logical jumps in information retrieval by relational indexing. M.Sc. thesis. London, The City University, 1970.

85. SALTON, G. Information storage and retrieval. Scientific Report No. ISR-8. Cambridge, Mass., Harvard University, 1964.

86. SALTON, G. The "generality" effect and the retrieval evaluation for large collections. J. Amer. Soc. Inf. Sci., 1972, 23, 11-22.

87. SARACEVIC, T. Selected results from an inquiry into testing of information retrieval systems. J. Amer. Soc. Inf. Sci., 1971, 22, 126-39.

88. SARACEVIC, T. and REES, A.M. Towards the identification and control of variables in information retrieval experimentation. J. Doc., 1967, 23, 7-19.

89. SCHEFFLER, F.L. and MARCH, J.F. Evaluation of the selective dissemination of information (SDI) program for the Aerospace Materials Information Center. Report AFML-TR-71-11, Ohio, Wright and Patterson AFB, 1971.

90. SCHNEIDER, J.H. Selective dissemination and indexing of information. Science, 1971, 173, 300-8.

91. SELYE, H. and EMBER, G. Symbolic Shorthand System (SSS) for physiology and medicine. Montreal, I.M.C.E., 1964.

92. SNYDER, M.B., SCHUMACHER, A.W., MAYER, S.E. and HARRON, M.D. Methodology for test and evaluation of document retrieval systems: a critical review and recommendations. Report HSR-RR-66/6-8k. McLean, Va., Human Sciences Research Inc., 1966. (PB-169 572)

93. SPARCK JONES, K. A statistical interpretation of term specificity and its application in retrieval. J. Doc., 1972, 28, 11-21.

94. SVENONIUS, E. An experiment in index term frequency. J. Amer. Soc. Inf. Sci., 1972, 23, 109-21.

95. SWETS, J.A. Information retrieval systems. Science, 1963, 141, 245-50.

96. TE NUYL, Th.W. The "l'unite" system with multiple search facility in the form of a Texoler sorter. Proc. Aslib 38th Annual Conference, Exeter, 1964, Looking forward in documentation, London, Aslib, 1964, 2-29 - 2-37.

97. UKCIS. Search Manual, 1971 edition. Nottingham, UKCIS, 1971.

98. VERHOEFF, J., GOFFMAN, W. and BELZER, J. Inefficiency of the use of Boolean functions. Comm. Assoc. Comp. Mach., 1961, 4, 557-8, 594.

99. WAGNER, R.H. A selective current awareness system using Engineering Index's Plastics data base. II. Performance. J. chem. Doc., 1969, 9, 85-8.

100. WOLFE, M.A. and HERNER, S. SDI system for the US Public Health Service, Office of Pesticides. J. chem. Doc., 1967. 7, 138-41.

101. YATES-MERCER, P.A. Logical jumps in information retrieval by relational indexing. M.Sc. thesis. London, The City University, 1969.

102. ZUNDE, P. and DEXTER, M.E. Indexing consistency and quality. Amer. Doc., 1969, 20, 259-67.

THE CITY UNIVERSITY

St. John Street, London, E.C.1

Telephone: CLErkenwell 4399

CENTRE FOR INFORMATION SCIENCE

From: Miss P.A. Yates-Mercer Reference: PAYM/sjl/ /

To: Date:

Relational Indexing applied to an S.D.I. System

Would you please scan sections 1.2, 1.4, 3.2, 3.3 and 3.4 (or those agreed upon when constructing your interest profile) of the issue of Metals Abstracts, and enter the reference numbers of those abstracts of interest to you in one or other of the columns below. Please return this form to me at the above address as soon as possible.

Abstracts of definite interest (i.e. papers that you would want to read)	Abstracts of possible interest (i.e. abstracts of interest, but papers that you would not necessarily want or have time to read)

N.B. If you have already seen the original paper and it falls within either of these categories, please include it in your list.

Comments (if any)

APPENDIX I

APPENDIX II

Tables of results

(arranged by set)

Set 071

Profile	Recall	Precision	Fallout	Profile	Recall	Precision	Fallout
1	0.0000	0.0000	0.0028	40	0.0000	0.0000	0.0181
3	1.0000	1.0000	0.0000	41	0.1250	0.2500	0.0042
6	0/0	0/0	0.0000	42	0.1111	0.1111	0.0112
7	0.0833	0.1250	0.0099	43	0.6667	0.3333	0.0056
9	0.7500	0.4286	0.0056	47	0.5882	0.8333	0.0028
10	0.1667	0.1429	0.0084	51	0.0000	0.0000	0.0014
12	0.6324	0.8269	0.0137	53	0.3214	0.4737	0.0144
13	0.2000	0.1739	0.0270	54	0.1111	0.0870	0.0298
14	0.0000	0.0000	0.0014	56	0.3000	0.1200	0.0309
21	0.2727	0.4286	0.0056	57	0.1818	0.1667	0.0140
22	0.3636	0.4444	0.0070	58	0.3333	0.0526	0.0250
24	1.0000	0.6429	0.0142	59	0.4500	0.3000	0.0299
26	0.3846	0.1163	0.0535	60	0.2174	0.1724	0.0343
27	0.2963	0.1905	0.0489	61	0.5000	0.1429	0.0083
29	0.6216	0.5897	0.0233	62	0.6250	0.1220	0.0504
30	0.0000	0.0000	0.0111	63	0.6250	0.5000	0.0070
31	0.5000	0.2308	0.0140	64	0.1177	0.1429	0.0170
32	0/0	0.0000	0.0000	65	0.2857	0.0952	0.0265
34	0.0333	0.1000	0.0130	66	0.5000	0.1364	0.0535
35	1.0000	0.8000	0.0014	67	1.0000	0.1667	0.0069
36	0.4286	1.0000	0.0000	69	0.5000	0.6250	0.0042
37	0.4000	0.3077	0.0126				

Set 072

Profile	Recall	Precision	Fallout	Profile	Recall	Precision	Fallout
1	0.0000	0.0000	0.0028	40	0.2000	0.0769	0.0167
3	1.0000	1.0000	0.0000	41	0.1429	0.2500	0.0042
6	0/0	0/0	0.0000	42	0.4167	0.5556	0.0056
7	0.1667	0.2500	0.0084	43	0.7500	0.5000	0.0042
9	0.7500	0.4286	0.0056	47	0.6316	1.0000	0.0000
10	0.3750	0.4286	0.0056	51	0.0000	0.0000	0.0014
12	0.6711	0.9808	0.0016	53	0.3448	0.5263	0.0130
13	0.2000	0.1739	0.0270	54	0.2382	0.2174	0.0256
14	0.0000	0.0000	0.0014	56	0.3000	0.1200	0.0309
21	0.5000	0.5714	0.0042	57	0.1818	0.1667	0.0140
22	0.5625	1.0000	0.0000	58	0.5000	0.0526	0.0250
24	1.0000	0.8571	0.0057	59	0.5500	0.3667	0.0270
26	0.6522	0.3488	0.0400	60	0.4849	0.5517	0.0188
27	0.3571	0.2381	0.0460	61	0.8000	0.5714	0.0042
29	0.8044	0.9487	0.0030	62	0.9143	0.7805	0.0131
30	0.1667	0.1250	0.0098	63	0.7273	0.8000	0.0028
31	0.7500	0.6923	0.0056	64	0.1250	0.1429	0.0170
32	1.0000	0.1667	0.0069	65	0.2857	0.0952	0.0265
34	0.0645	0.2000	0.0116	66	0.5000	0.1364	0.0535
35	1.0000	0.8000	0.0014	67	1.0000	0.3333	0.0056
36	0.4286	1.0000	0.0000	69	0.5000	0.6250	0.0042
37	0.6429	0.6923	0.0056				

Set 073

Profile	Recall	Precision	Fallout	Profile	Recall	Precision	Fallout
1	0.0000	0.0000	0.0014	40	0.2000	0.2500	0.0042
3	1.0000	1.0000	0.0000	41	0.2857	0.5000	0.0028
6	0/0	0/0	0.0000	42	0.9167	0.7333	0.0056
7	0.5833	0.7778	0.0028	43	1.0000	0.7500	0.0014
9	0.7500	0.4286	0.0056	47	0.6842	1.0000	0.0000
10	0.8750	1.0000	0.0000	51	0/0	0/0	0.0000
12	0.8947	1.0000	0.0000	53	0.5172	0.7895	0.0058
13	0.6000	0.4737	0.0141	54	0.6316	0.5000	0.0171
14	0.7500	1.0000	0.0000	56	0.6667	0.7500	0.0028
21	0.7500	0.8571	0.0014	57	0.6667	0.3750	0.0140
22	0.8125	1.0000	0.0000	58	0.5000	0.5000	0.0014
24	1.0000	0.8571	0.0057	59	0.9231	0.6667	0.0085
26	0.7826	0.8571	0.0043	60	0.6061	0.7407	0.0101
27	0.7727	0.8095	0.0057	61	1.0000	1.0000	0.0000
29	0.8696	1.0000	0.0000	62	1.0000	0.8250	0.0101
30	0.1667	0.5000	0.0014	63	0.7273	0.8000	0.0028
31	0.7500	0.9000	0.0014	64	0.3750	0.4615	0.0099
32	1.0000	0.5000	0.0014	65	0.2857	0.1333	0.0182
34	0.1250	0.6667	0.0014	66	0.8182	1.0000	0.0000
35	1.0000	0.8000	0.0014	67	1.0000	0.4000	0.0042
36	0.4286	1.0000	0.0000	69	0.7000	1.0000	0.0000
37	0.8571	1.0000	0.0000				

Set 083

Profile	Recall	Precision	Fallout	Profile	Recall	Precision	Fallout
1	0/0	0/0	0.0000	40	0.7500	0.7059	0.0116
3	1.0000	1.0000	0.0000	41	0.2222	1.0000	0.0000
6	0/0	0/0	0.0000	42	0.5000	0.5000	0.0045
7	0.7778	0.5833	0.0114	43	0.6667	0.6667	0.0023
9	0.0000	0/0	0.0000	47	0.9231	1.0000	0.0000
10	0.8333	0.7143	0.0045	51	1.0000	0.6667	0.0022
12	0.8529	1.0000	0.0000	53	0.5682	0.9615	0.0025
13	0.0000	0/0	0.0000	54	0.6000	0.7500	0.0023
14	1.0000	0.5000	0.0067	56	1.0000	0.7500	0.0023
21	1.0000	1.0000	0.0000	57	0.6000	0.6000	0.0091
22	0.5000	0.2000	0.0090	58	1.0000	0.5000	0.0022
24	1.0000	1.0000	0.0000	59	0.6667	1.0000	0.0000
26	0.6875	0.7857	0.0069	60	0.9286	0.7647	0.0092
27	0.5000	0.6667	0.0069	61	1.0000	1.0000	0.0000
29	0.6829	0.9333	0.0049	62	0.5000	0.4444	0.0114
30	0.6250	0.5556	0.0091	63	1.0000	0.6667	0.0022
31	0.5000	0.6667	0.0023	64	0.6000	0.2727	0.0181
32	1.0000	0.5000	0.0023	65	1.0000	0.8000	0.0023
34	0.6000	0.3156	0.0300	66	0.4000	0.3333	0.0090
35	0/0	0/0	0.0000	67	0/0	0/0	0.0000
36	0.7500	1.0000	0.0000	69	0.6000	0.5000	0.0068
37	1.0000	1.0000	0.0000				

Set 093

Profile	Recall	Precision	Fallout	Profile	Recall	Precision	Fallout
1	0/0	0/0	0.0000	40	0.0000	0/0	0.0000
3	0/0	0/0	0.0000	41	0.8571	1.0000	0.0000
6	0/0	0/0	0.0000	42	0.8000	1.0000	0.0000
7	1.0000	0.5000	0.0140	43	1.0000	1.0000	0.0000
9	0.5000	1.0000	0.0000	47	1.0000	1.0000	0.0000
10	0.5000	0.5714	0.0107	51	0.6667	0.4000	0.0105
12	0.9286	1.0000	0.0000	53	0.7500	0.8000	0.0110
13	1.0000	0.5000	0.0035	54	0.8889	0.6154	0.0179
14	0.8333	0.8333	0.0035	56	0.5000	1.0000	0.0000
21	1.0000	1.0000	0.0000	57	0.3333	1.0000	0.0000
22	0.2500	1.0000	0.0000	58	1.0000	1.0000	0.0000
24	0.9286	1.0000	0.0000	59	1.0000	0.8333	0.0035
26	0.6250	0.8333	0.0036	60	0.8333	0.9091	0.0036
27	0.5000	1.0000	0.0000	61	1.0000	1.0000	0.0000
29	0.8696	0.8889	0.0206	62	1.0000	0.8750	0.0036
30	0.5000	1.0000	0.0000	63	1.0000	1.0000	0.0000
31	1.0000	1.0000	0.0000	64	1.0000	1.0000	0.0000
32	1.0000	1.0000	0.0000	65	1.0000	0.7500	0.0035
34	0.4118	1.0000	0.0000	66	0.8000	0.5714	0.0106
35	0/0	0/0	0.0000	67	0/0	0/0	0.0000
36	0.6000	1.0000	0.0000	69	0.6667	0.6667	0.0035
37	1.0000	0.7500	0.0035				

Set 103

Profile	Recall	Precision	Fallout	Profile	Recall	Precision	Fallout
1	0/0	0/0	0.0000	40	0.3333	1.0000	0.0000
3	1.0000	1.0000	0.0000	41	0.5000	0.3333	0.0068
6	0/0	0/0	0.0000	42	0.6667	1.0000	0.0000
7	0.6667	1.0000	0.0000	43	1.0000	1.0000	0.0000
9	1.0000	0.5000	0.0034	47	0.6667	1.0000	0.0000
10	1.0000	1.0000	0.0000	51	0.7500	0.7500	0.0034
12	0.9524	1.0000	0.0000	53	0.7059	0.9231	0.0036
13	1.0000	1.0000	0.0000	54	1.0000	0.7500	0.0034
14	0.6667	0.6667	0.0034	56	1.0000	1.0000	0.0000
21	0.8000	0.8000	0.0034	57	0.5000	1.0000	0.0000
22	1.0000	1.0000	0.0000	58	0.6000	0.7500	0.0034
24	1.0000	0.8462	0.0070	59	1.0000	0.8000	0.0034
26	0.5294	0.7500	0.0107	60	0.5000	0.7143	0.0069
27	0.5455	0.8571	0.0035	61	1.0000	1.0000	0.0000
29	0.6191	0.8667	0.0072	62	0.9091	0.9091	0.0035
30	0.5000	0.5000	0.0034	63	1.0000	1.0000	0.0000
31	0.8000	1.0000	0.0000	64	0.7500	0.7500	0.0034
32	1.0000	0.7500	0.0034	65	0.8000	0.6667	0.0068
34	0.7083	0.8500	0.0110	66	0.8571	0.6667	0.0103
35	0/0	0/0	0.0000	67	1.0000	1.0000	0.0000
36	0.6667	1.0000	0.0000	69	0.6667	0.5000	0.0068
37	0.8000	1.0000	0.0000				

Set 113

Profile	Recall	Precision	Fallout	Profile	Recall	Precision	Fallout
1	1.0000	1.0000	0.0000	40	0.8571	0.8571	0.0017
3	1.0000	0.5000	0.0017	41	0.4000	1.0000	0.0000
6	0/0	0/0	0.0000	42	0.9000	0.9000	0.0017
7	0.8000	0.6667	0.0034	43	1.0000	0.8333	0.0017
9	0.8333	0.6250	0.0052	47	0.9500	0.7600	0.0106
10	0.9000	0.6000	0.0104	51	1.0000	1.0000	0.0000
12	1.0000	1.0000	0.0000	53	0.7778	0.8235	0.0053
13	0.8000	0.5714	0.0052	54	0.8333	0.4167	0.0121
14	1.0000	0.6667	0.0017	56	1.0000	0.8571	0.0017
21	0.8889	0.7273	0.0052	57	0.5455	0.4615	0.0122
22	1.0000	0.3333	0.0068	58	0.5000	1.0000	0.0000
24	0.9546	0.9546	0.0018	59	0.7143	1.0000	0.0000
26	0.7273	0.6667	0.0069	60	0.3462	0.6429	0.0089
27	0.4546	0.4167	0.0122	61	1.0000	0.8182	0.0035
29	0.7714	0.8710	0.0073	62	0.9677	1.0000	0.0000
30	0.8333	0.8333	0.0017	63	0.7727	1.0000	0.0000
31	1.0000	1.0000	0.0000	64	0.6000	0.3750	0.0086
32	0.3333	1.0000	0.0000	65	0.3000	0.3750	0.0087
34	0.8710	0.8710	0.0072	66	0.9167	0.6875	0.0087
35	0.7500	1.0000	0.0000	67	1.0000	0.2727	0.0137
36	0.8333	1.0000	0.0000	69	0.7727	0.8947	0.0035
37	0.7333	0.7857	0.0052				

Set 121

Profile	Recall	Precision	Fallout	Profile	Recall	Precision	Fallout
1	0/0	0.0000	0.0063	40	0.6250	0.5000	0.0106
3	0.3333	1.0000	0.0000	41	0.4286	0.3333	0.0127
6	0/0	0/0	0.0000	42	0.7778	0.7000	0.0064
7	0.5714	0.3636	0.0148	43	0.6667	0.3333	0.0084
9	0.8182	0.7500	0.0064	47	0.5714	0.6667	0.0042
10	0.5000	0.2857	0.0105	51	0.0000	0.0000	0.0105
12	0.9583	0.7419	0.0175	53	0.5333	0.3810	0.0280
13	0.6667	1.0000	0.0000	54	0.5000	0.3750	0.0106
14	0.0000	0.0000	0.0042	56	0.6667	1.0000	0.0000
21	0.4000	0.2000	0.0168	57	0.3500	0.7000	0.0065
22	0.1667	0.1667	0.0106	58	0.4000	1.0000	0.0000
24	0.8947	0.9444	0.0022	59	0.8000	0.6667	0.0042
26	0.6667	0.1379	0.0527	60	0.0968	0.1579	0.0356
27	0.3529	0.4286	0.0173	61	0.0000	0.0000	0.0146
29	0.7895	0.5172	0.0304	62	0.2500	0.3000	0.0150
30	0.3333	0.2000	0.0084	63	0.5000	1.0000	0.0000
31	0.6154	0.6154	0.0107	64	0.0000	0.0000	0.0126
32	0.5000	0.2500	0.0126	65	0.1000	0.1111	0.0170
34	0.1667	0.0303	0.0675	66	0.5385	0.7000	0.0064
35	0/0	0/0	0.0000	67	0/0	0/0	0.0000
36	1.0000	0.3000	0.0147	69	0.6364	0.6667	0.0153
37	0.7143	0.4167	0.0761				

Set 122

Profile	Recall	Precision	Fallout	Profile	Recall	Precision	Fallout
1	1.0000	1.0000	0.0000	40	0.8182	0.9000	0.0021
3	0.3333	1.0000	0.0000	41	0.6000	0.6667	0.0064
6	0/0	0/0	0.0000	42	0.8750	0.7000	0.0064
7	0.7000	0.6364	0.0085	43	0.8571	1.0000	0.0000
9	0.8182	0.7500	0.0064	47	0.6667	0.6667	0.0042
10	0.7143	0.7143	0.0042	51	0.5000	0.4000	0.0063
12	0.9677	0.9677	0.0022	53	0.6500	0.6191	0.0174
13	0.6667	1.0000	0.0000	54	0.5714	0.5000	0.0085
14	0.2500	0.5000	0.0021	56	0.6667	1.0000	0.0000
21	0.7000	0.7000	0.0064	57	0.3810	0.8000	0.0044
22	0.2000	0.1667	0.0105	58	0.6667	1.0000	0.0000
24	0.8947	0.9444	0.0022	59	0.8333	0.8333	0.0021
26	0.8947	0.5862	0.0260	60	0.3333	0.7368	0.0114
27	0.4286	0.4286	0.0172	61	0.7500	0.4286	0.0084
29	0.8667	0.8966	0.0067	62	0.5333	0.8000	0.0043
30	0.3333	0.2000	0.0084	63	0.5000	1.0000	0.0000
31	0.6667	0.7692	0.0065	64	0.6667	0.3333	0.0084
32	0.7500	0.7500	0.0042	65	0.3077	0.4444	0.0107
34	0.8519	0.6970	0.0221	66	0.6000	0.9000	0.0022
35	0/0	0/0	0.0000	67	0/0	0/0	0.0000
36	1.0000	1.0000	0.0000	69	0.6957	0.7619	0.0109
37	0.8571	1.0000	0.0000				

Set 123

Profile	Recall	Precision	Fallout	Profile	Recall	Precision	Fallout
1	1.0000	1.0000	0.0000	40	0.8182	1.0000	0.0000
3	0.5000	1.0000	0.0000	41	0.6667	1.0000	0.0000
6	0/0	0/0	0.0000	42	0.8750	0.8750	0.0021
7	0.8750	0.7000	0.0064	43	0.8571	1.0000	0.0000
9	0.8182	0.9000	0.0021	47	0.6667	1.0000	0.0000
10	0.7143	0.7143	0.0042	51	1.0000	1.0000	0.0000
12	1.0000	0.9677	0.0022	53	0.7778	0.7778	0.0087
13	0.6667	1.0000	0.0000	54	0.8571	0.7500	0.0042
14	1.0000	1.0000	0.0000	56	0.8000	1.0000	0.0000
21	0.8889	1.0000	0.0000	57	0.4000	0.8889	0.0022
22	0.2000	0.5000	0.0021	58	1.0000	1.0000	0.0000
24	1.0000	0.9444	0.0022	59	1.0000	0.8333	0.0021
26	0.9474	0.6923	0.0174	60	0.5385	1.0000	0.0000
27	0.8182	0.6923	0.0085	61	0.7500	0.7500	0.0021
29	0.8667	0.8966	0.0067	62	0.6667	0.8889	0.0021
30	1.0000	1.0000	0.0000	63	0.7500	1.0000	0.0000
31	0.7692	0.7692	0.0064	64	1.0000	0.5000	0.0042
32	1.0000	1.0000	0.0000	65	0.3333	0.5000	0.0086
34	0.9200	0.9200	0.0044	66	1.0000	1.0000	0.0000
35	0/0	0/0	0.0000	67	0/0	0/0	0.0000
36	1.0000	1.0000	0.0000	69	0.8947	0.7727	0.0109
37	1.0000	1.0000	0.0000				

Set 123

Profile	G	G' (with second value of N)	F'	Profile	G	G' (with second value of N)	F'
1	0.0063	0.0063	0.0000	40	0.0229	0.0229	0.0000
3	0.0042	0.0042	0.0000	41	0.0188	0.0188	0.0000
6	0.0000	0.0000	0.0000	42	0.0167	0.0167	0.0021
7	0.0167	0.0167	0.0064	43	0.0146	0.0707	0.0000
9	0.0229	0.0229	0.0021	47	0.0125	0.0125	0.0000
10	0.0146	0.0146	0.0042	51	0.0083	0.0083	0.0000
12	0.0625	0.0625	0.0022	53	0.0375	0.0375	0.0087
13	0.0125	0.0125	0.0000	54	0.0146	0.0146	0.0042
14	0.0021	0.0021	0.0000	56	0.0104	0.0104	0.0000
21	0.0188	0.0188	0.0000	57	0.0417	0.0417	0.0022
22	0.0104	0.0104	0.0021	58	0.0042	0.0042	0.0000
24	0.0354	0.0354	0.0022	59	0.0104	0.0104	0.0021
26	0.0396	0.0396	0.0174	60	0.0813	0.1024	0.0000
27	0.0229	0.0229	0.0085	61	0.0083	0.0404	0.0105
29	0.0625	0.0625	0.0067	62	0.0250	0.0250	0.0021
30	0.0042	0.0042	0.0000	63	0.0167	0.0167	0.0000
31	0.0271	0.0271	0.0064	64	0.0042	0.0042	0.0042
32	0.0125	0.0125	0.0000	65	0.0250	0.0678	0.0242
34	0.0521	0.0521	0.0044	66	0.0229	0.0534	0.0000
35	0.0000	0.0000	0.0000	67	0.0000	0.0000	0.0000
36	0.0208	0.0208	0.0000	69	0.0396	0.2375	0.0820
37	0.0292	0.1414	0.0000				

Set 123

Profile	$\frac{1}{M}$	$\frac{1}{G}$	$\frac{1}{P}$	$\frac{1}{F}$	Profile	$\frac{1}{M}$	$\frac{1}{G}$	$\frac{1}{P}$	$\frac{1}{F}$
1	1.00	159	1.00	I	40	1.22	44	1.00	I
3	2.00	238	1.00	I	41	1.50	53	1.00	I
6	-	-	-	I	42	1.14	60	1.14	476
7	1.14	60	1.43	156	43	1.17	68	1.00	I
9	1.22	44	1.11	476	47	1.50	80	1.00	I
10	1.40	68	1.40	238	51	1.00	121	1.00	I
12	1.00	16	1.03	455	53	1.28	27	1.29	115
13	1.50	80	1.00	I	54	1.17	68	1.33	238
14	1.00	476	1.00	I	56	1.25	96	1.00	I
21	1.12	53	1.00	I	57	2.50	24	1.12	455
22	5.00	96	2.00	476	58	1.00	238	1.00	I
24	1.00	28	1.06	455	59	1.00	96	1.20	476
26	1.06	25	1.44	57	60	1.86	12	1.00	I
27	1.22	44	1.44	12	61	1.33	121	1.33	476
29	1.15	16	1.12	149	62	1.50	40	1.12	476
30	1.00	238	1.00	I	63	1.33	60	1.00	I
31	1.30	37	1.30	156	64	1.00	238	2.00	238
32	1.00	80	1.00	I	65	3.00	40	2.00	116
34	1.09	19	1.09	227	66	1.00	44	1.00	I
35	-	-	-	I	67	-	-	-	I
36	1.00	48	1.00	I	69	1.12	25	1.29	92
37	1.00	34	1.00	I					

'I' represents infinity

Set 123 (Relevance 1)

Profile	Recall	Precision	Fallout	Profile	Recall	Precision	Fallout
1	0/0	0/0	0.0000	40	0.2500	1.0000	0.0000
3	0.0000	0/0	0.0000	41	0.0000	0.0000	0.0021
6	0/0	0/0	0.0000	42	0.0000	0.0000	0.0042
7	0.0000	0.0000	0.0063	43	0/0	0.0000	0.0042
9	0.0000	0.0000	0.0021	47	1.0000	1.0000	0.0000
10	1.0000	0.5000	0.0021	51	0/0	0/0	0.0000
12	0.5714	0.8421	0.0066	53	0.2000	0.5000	0.0043
13	1.0000	1.0000	0.0000	54	0.3333	1.0000	0.0000
14	0/0	0/0	0.0000	56	0.2500	1.0000	0.0000
21	0.0000	0/0	0.0000	57	0.4000	1.0000	0.0000
22	0.0000	0/0	0.0000	58	0/0	0.0000	0.0042
24	0.5000	0.1000	0.0188	59	0.2500	1.0000	0.0000
26	0.0000	0.0000	0.0021	60	0.1111	1.0000	0.0000
27	0.6667	0.5000	0.0042	61	0.0000	0/0	0.0000
29	0.0000	0.0000	0.0042	62	0.3333	0.2000	0.0084
30	0.0000	0/0	0.0000	63	0.0000	0/0	0.0000
31	0.6000	0.4286	0.0084	64	0.0000	0/0	0.0000
32	0.5000	1.0000	0.0000	65	0.4000	0.5000	0.0042
34	0.5455	0.4286	0.0171	66	0.0000	0.0000	0.0105
35	0/0	0/0	0.0000	67	0/0	0/0	0.0000
36	0/0	0.0000	0.0083	69	0.7500	0.3750	0.0105
37	0.5000	0.2857	0.0105				

Set 123 (Relevance 2)

Profile	Recall	Precision	Fallout	Profile	Recall	Precision	Fallout
1	1.0000	1.0000	0.0000	40	0.7143	0.6250	0.0063
3	0.0000	0.0000	0.0021	41	0.6667	0.8000	0.0021
6	0/0	0/0	0.0000	42	0.6000	0.5000	0.0063
7	0.2000	0.1429	0.0126	43	0.5714	1.0000	0.0000
9	0.7000	0.7778	0.0043	47	0.5000	1.0000	0.0000
10	0.5000	0.6000	0.0042	51	1.0000	1.0000	0.0000
12	0.0000	0.0000	0.0251	53	0.6250	0.3571	0.0191
13	0.3333	1.0000	0.0000	54	1.0000	0.5714	0.0063
14	1.0000	1.0000	0.0000	56	0.0000	0.0000	0.0063
21	1.0000	0.3750	0.0105	57	0.3333	0.7143	0.0043
22	0.2500	0.5000	0.0021	58	0.0000	0/0	0.0000
24	0.4667	0.8750	0.0022	59	1.0000	0.2000	0.0084
26	1.0000	0.4400	0.0299	60	0.7143	0.7895	0.0087
27	0.6250	0.5556	0.0085	61	1.0000	0.5000	0.0042
29	0.8519	0.8519	0.0088	62	0.3333	0.7500	0.0021
30	0/0	0.0000	0.0042	63	0.8333	0.8333	0.0021
31	0.1250	0.1667	0.0106	64	0/0	0.0000	0.0083
32	1.0000	0.5000	0.0042	65	0.0000	0.0000	0.0085
34	0.4286	0.5455	0.0107	66	0.3750	0.5000	0.0064
35	0/0	0/0	0.0000	67	0/0	0/0	0.0000
36	0.6000	1.0000	0.0000	69	0.5333	0.5714	0.0129
37	0.5000	0.7143	0.0043				

Set 708

Profile	Recall	Precision	Fallout	Profile	Recall	Precision	Fallout
1	0.0000	0.0000	0.0009	40	0.6191	0.6191	0.0070
3	1.0000	1.0000	0.0000	41	0.2500	0.6667	0.0017
6	0/0	0/0	0.0000	42	0.8125	0.6842	0.0052
7	0.6667	0.6667	0.0061	43	0.8333	0.7143	0.0017
9	0.5000	0.4286	0.0034	47	0.7813	1.0000	0.0000
10	0.8571	0.8571	0.0017	51	1.0000	0.5000	0.0017
12	0.8818	1.0000	0.0000	53	0.5480	0.8889	0.0046
13	0.3600	0.4737	0.0087	54	0.6250	0.5357	0.0113
14	0.8571	0.6667	0.0026	56	0.7500	0.7500	0.0026
21	0.8182	0.9000	0.0009	57	0.6316	0.4615	0.0122
22	0.7778	0.7778	0.0035	58	0.6667	0.5000	0.0017
24	1.0000	0.9111	0.0035	59	0.8421	0.7273	0.0052
26	0.7436	0.8286	0.0053	60	0.7021	0.7500	0.0098
27	0.6765	0.7667	0.0062	61	1.0000	1.0000	0.0000
29	0.7816	0.9714	0.0019	62	0.9024	0.7551	0.0106
30	0.4286	0.6000	0.0035	63	0.7692	0.7692	0.0026
31	0.6875	0.8462	0.0017	64	0.4286	0.3750	0.0130
32	1.0000	0.5000	0.0017	65	0.5455	0.3000	0.0121
34	0.3077	0.3636	0.0122	66	0.6875	0.7333	0.0035
35	1.0000	0.8000	0.0009	67	1.0000	0.4000	0.0026
36	0.5455	1.0000	0.0000	69	0.6667	0.7692	0.0026
37	0.8667	1.0000	0.0000				

Set 709

Profile	Recall	Precision	Fallout	Profile	Recall	Precision	Fallout
1	0.0000	0.0000	0.0007	40	0.5652	0.6191	0.0056
3	1.0000	1.0000	0.0000	41	0.4348	0.8333	0.0014
6	0/0	0/0	0.0000	42	0.8095	0.7391	0.0042
7	0.7200	0.6207	0.0077	43	0.8750	0.7778	0.0014
9	0.5000	0.5000	0.0028	47	0.7879	1.0000	0.0000
10	0.7273	0.7619	0.0035	51	0.8000	0.4444	0.0034
12	0.8913	1.0000	0.0000	53	0.5843	0.8667	0.0058
13	0.3846	0.4762	0.0077	54	0.6970	0.5610	0.0126
14	0.8462	0.7333	0.0028	56	0.7143	0.7692	0.0021
21	0.8571	0.9231	0.0007	57	0.5909	0.4815	0.0097
22	0.6818	0.7895	0.0028	58	0.7500	0.6000	0.0014
24	0.9818	0.9310	0.0029	59	0.8750	0.7500	0.0049
26	0.7234	0.8293	0.0050	60	0.7288	0.7818	0.0086
27	0.6304	0.8056	0.0050	61	1.0000	1.0000	0.0000
29	0.8120	0.9391	0.0053	62	0.9167	0.7719	0.0092
30	0.4375	0.6364	0.0028	63	0.7857	0.7857	0.0021
31	0.7727	0.8947	0.0014	64	0.5000	0.4444	0.0105
32	1.0000	0.6667	0.0014	65	0.6429	0.3750	0.0104
34	0.3488	0.5556	0.0085	66	0.7143	0.6818	0.0049
35	1.0000	0.8000	0.0007	67	1.0000	0.4000	0.0021
36	0.5625	1.0000	0.0000	69	0.6667	0.7500	0.0028
37	0.8889	0.9412	0.0007				

Set 710

Profile	Recall	Precision	Fallout	Profile	Recall	Precision	Fallout
1	0.0000	0.0000	0.0006	40	0.5385	0.6364	0.0046
3	1.0000	1.0000	0.0000	41	0.4400	0.7333	0.0023
6	0/0	0/0	0.0000	42	0.7917	0.7600	0.0035
7	0.7143	0.6452	0.0064	43	0.8889	0.8000	0.0011
9	0.5556	0.5000	0.0029	47	0.7692	1.0000	0.0000
10	0.7692	0.8000	0.0029	51	0.7778	0.5385	0.0034
12	0.8994	1.0000	0.0000	53	0.6038	0.8767	0.0055
13	0.4483	0.5417	0.0064	54	0.7222	0.5778	0.0110
14	0.8125	0.7222	0.0029	56	0.8182	0.8571	0.0017
21	0.8421	0.8889	0.0012	57	0.5769	0.5172	0.0081
22	0.7083	0.8095	0.0023	58	0.6667	0.6667	0.0017
24	0.9849	0.9155	0.0036	59	0.8929	0.7576	0.0046
26	0.6719	0.8113	0.0059	60	0.6957	0.7742	0.0083
27	0.6140	0.8140	0.0047	61	1.0000	1.0000	0.0000
29	0.7857	0.9308	0.0056	62	0.9153	0.7941	0.0082
30	0.4444	0.6154	0.0029	63	0.8235	0.8235	0.0017
31	0.7813	0.9259	0.0012	64	0.5357	0.4839	0.0093
32	1.0000	0.7000	0.0017	65	0.6842	0.4333	0.0098
34	0.4776	0.6809	0.0089	66	0.7500	0.6774	0.0058
35	1.0000	0.8000	0.0006	67	1.0000	0.5000	0.0017
36	0.5790	1.0000	0.0000	69	0.6667	0.7000	0.0035
37	0.8696	0.9524	0.0006				

Set 711

Profile	Recall	Precision	Fallout	Profile	Recall	Precision	Fallout
1	0.2500	0.5000	0.0004	40	0.6061	0.6897	0.0039
3	1.0000	0.8750	0.0004	41	0.4333	0.7647	0.0017
6	0/0	0/0	0.0000	42	0.8235	0.8000	0.0030
7	0.7273	0.6487	0.0056	43	0.9286	0.8125	0.0013
9	0.6667	0.5556	0.0034	47	0.8305	0.8909	0.0026
10	0.8056	0.7250	0.0048	51	0.8182	0.6000	0.0026
12	0.9252	1.0000	0.0000	53	0.6290	0.8667	0.0054
13	0.5000	0.5484	0.0061	54	0.7381	0.5439	0.0113
14	0.8333	0.7143	0.0026	56	0.8571	0.8571	0.0017
21	0.8571	0.8276	0.0022	57	0.5676	0.5000	0.0091
22	0.7308	0.7037	0.0035	58	0.6364	0.7000	0.0013
24	0.9773	0.9247	0.0031	59	0.8571	0.7895	0.0035
26	0.6800	0.7846	0.0062	60	0.6000	0.7500	0.0084
27	0.5882	0.7273	0.0066	61	1.0000	0.9231	0.0009
29	0.7831	0.9193	0.0060	62	0.9333	0.8571	0.0062
30	0.5417	0.6842	0.0026	63	0.7949	0.9118	0.0013
31	0.8108	0.9375	0.0009	64	0.5455	0.4615	0.0091
32	0.8000	0.7273	0.0013	65	0.5517	0.4211	0.0095
34	0.6020	0.7564	0.0085	66	0.8000	0.6809	0.0065
35	0.8750	0.8750	0.0004	67	1.0000	0.3529	0.0047
36	0.6774	1.0000	0.0000	69	0.7209	0.7949	0.0035
37	0.8158	0.8857	0.0017				

Set 712

Profile	M	P	F	G	Profile	M	P	F	G
1	0.5714	0.8000	0.0004	0.0025	40	0.6591	0.7632	0.0032	0.0156
3	0.8889	0.8889	0.0004	0.0032	41	0.4872	0.8261	0.0014	0.0138
6	0/0	0/0	0.0000	0.0000	42	0.8333	0.8140	0.0029	0.0149
7	0.7561	0.6596	0.0058	0.0145	43	0.9048	0.8636	0.0011	0.0074
9	0.7308	0.6786	0.0032	0.0092	47	0.8154	0.8983	0.0022	0.0230
10	0.7907	0.7234	0.0047	0.0152	51	0.8667	0.6842	0.0021	0.0053
12	0.9344	0.9956	0.0004	0.0864	53	0.6479	0.8519	0.0060	0.0503
13	0.5250	0.6000	0.0050	0.0142	54	0.7551	0.5692	0.0101	0.0174
14	0.8421	0.7273	0.0021	0.0067	56	0.8485	0.8750	0.0014	0.0117
21	0.8649	0.8649	0.0018	0.0131	57	0.5088	0.5686	0.0080	0.0202
22	0.6452	0.6897	0.0032	0.0110	58	0.6923	0.7500	0.0011	0.0046
24	0.9810	0.9279	0.0029	0.0372	59	0.8750	0.7955	0.0032	0.0142
26	0.7340	0.7582	0.0081	0.0333	60	0.5821	0.8041	0.0071	0.0474
27	0.6203	0.7206	0.0069	0.0280	61	0.9643	0.9000	0.0011	0.0099
29	0.7945	0.9158	0.0061	0.0775	62	0.9020	0.8598	0.0055	0.0361
30	0.5769	0.7143	0.0021	0.0092	63	0.7708	0.9250	0.0011	0.0170
31	0.8000	0.8889	0.0018	0.0177	64	0.5714	0.4651	0.0082	0.0124
32	0.8750	0.8235	0.0011	0.0057	65	0.4878	0.4348	0.0093	0.0145
34	0.6667	0.7961	0.0078	0.0435	66	0.8431	0.7414	0.0054	0.0181
35	0.8750	0.8750	0.0004	0.0028	67	1.0000	0.3529	0.0039	0.0021
36	0.7561	1.0000	0.0000	0.0145	69	0.7742	0.7869	0.0047	0.0220
37	0.8654	0.9184	0.0014	0.0184					

APPENDIX III

Examples of abstracts with their relationally indexed displays

1. 7012 12/1447

The crystal structure of a complex boride, NaB_{15}, was determined by X-ray diffraction methods using single crystals. The structure consists principally of B_{12} icosahedra in a novel way but with additional B atoms linking icosahedra together. The Na atoms occur in large cages that result from the B arrangement.

2. 7012 32/0774

The theory of the external two-quantum photo-effect in metals is presented and discussed with special reference to the question of polarization selectivity. The polarization selectivity of the two-quantum photo-effect differs from that of the one-quantum effect in having a more sharply defined region corresponding to the optimum spatial orientation of the electrical-field vector of the incident radiation for which the photocurrent increases rapidly. Differences hitherto noted between theory and experiment may have been due to a neglect of the vectorial character of the photo-effect.

3. 7012 12/1393

The distribution of I impurities in GaAs crystals was studied by a radioisotope method after growing the crystals from the melt by the Czochralski technique. The distribution of the I along the growing crystal was in excellent agreement with a curve calculated from the Pfann equation, with an effective distribution coefficient of 3.5×10^{-3}. Special orientational effects were exhibited by the I impurity for GaAs crystal grown in the <111> directions; an example is illustrated.

1.

NaB_{15} /+ single crystal /(structure /- measuring /; X-ray diffraction

2.

metals /; photo-effect [external two-quantum]
/* /;
polarization /; selectivity /θ theory /- discussing

3.

GaAs /+ crystal /; growing /; Czochralski method
/;
I /= impurity
/+
distribution /- studying /; radioisotope
/;
coefficient

4. 7012 33/1941

The effect of pressure on the Neel temperature of Cr-<3% V alloys was studied by electrical resistivity measurements under hydrostatic pressures up to 3000 atm. The Neel temperature varies linearly with pressure. The results are discussed in the light of recent theories.

5. 7010 34/0498

Cu^+ activity at the surface of non-polarized Cu electrodes in $CuSO_4$ solution was investigated by measuring redox potentials with a Pt micro-electrode. Cu^+ oxidation by dissolved O occurs at a certain distance below the surface of the Cu electrode. The reaction takes place in a 0.1 mm thick layer of solution in which the change of redox potential reaches 200-300 mV. The distance of the reaction region from the electrode surface is inversely proportional to the concentration of dissolved O. In the vicinity of the electrode surface a_{Cu^+} is practically independent on the concentration of dissolved O and approaches the equilibrium value.

6. 7012 14/0480

The ageing of high-coercivity alloys of the Alnico type (Magnico and Ticonal) was studied in relation to their ultimate magnetic characteristics. The single and multistage ageing conventionally employed failed to give the best combination of such properties as required for permanent magnets. These properties were in fact very sensitive to the precise programming of the heat treatment, and in general the repeated lowering of the temperature between stages tended to increase the coercive force and maximum magnetic energy.

4.

alloy /({Cr / V /+ < 3% /+ antiferromagnetic pressure, hydrostatic /= up to 3000 atm /; /; Neel temperature /- studying /; electrical conductivity

5.

Cu /+ electrode /(/; Cu^+ /+ surface /; activity /- studying /; potential, redox /- measuring /; Pt /+ electrode, micro- /; polarizing /+ $CuSO_4$ /+ solution

6.

alloy /({Al / Ni / Co_ magnetic properties /- /= Alnico /+ magnetic /; coercivity [high] /* a g e i n g /- studying relating

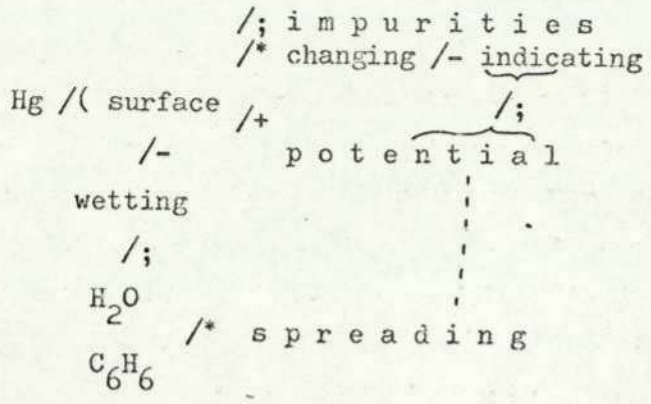
7. 7010 32/0650

A fresh surface of Hg undergoes a change with time that is indicated by the spreading of water and benzene on the surface and by the relative surface potential. After a period of time the surface appears to become static, as characterized by no further changes in the spreading characteristics of water and benzene or of the surface potential. These wetting and surface potential properties - both the change with time and the final equilibrium values - are greatly affected by the presence of trace amounts, such as 2 ppm of Zn and Pb; Au had no effect. The effect is almost immediate upon application of the contaminant. The changing surface properties appear to be due to the adsorption of gas molecules on the surface, for which the impurities act as a catalyst, or to the presence of impurity molecules in the surface.

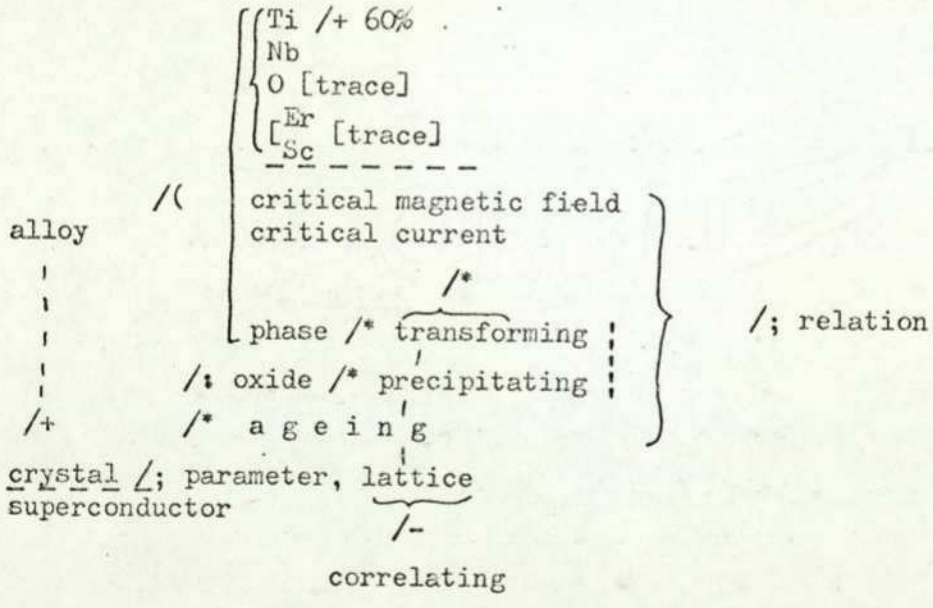
8. 7010 33/1655

The aging behaviour of a Nb-Ti alloy containing 60 wt.% Ti and small additions of O and Er or Sc was characterized and related to superconducting properties. The ternary and quaternary alloys were cold reduced and aged for various times at temperatures between 250 and 1000°C. The ω and α-phase transformations and oxide precipitation processes were followed by lattice parameter, diffraction intensity, resistivity and metallographic studies and correlated with superconducting critical magnetic field and critical current density measurements. The optimum 1-h aging temperature for producing ω and α-phase precipitation were found to be 400 and 500°C, respectively. Aging at 1000°C produced only oxide precipitation. It was found that the O, Er and Sc stabilize the α-phase but have little effect on ω precipitation. The ω phase proved the most effective fluxoid-pinning precipitate. The fine-scale dispersoid provided an extremely high number density of pinning sites.

7.


/; i m p u r i t i e s
/* changing /- indicating
Hg /(surface
/+
/;
/-
p o t e n t i a l
wetting
/;
H_2O
/* s p r e a d i n g
C_6H_6

8.


Ti /+ 60%
Nb
O [trace]
[Er / Sc [trace]
/(
critical magnetic field
critical current
/*
phase /* transforming
/; oxide /* precipitating
/* a g e i n g
alloy
/+
/; relation
crystal / superconductor
/; parameter, lattice
/-
correlating

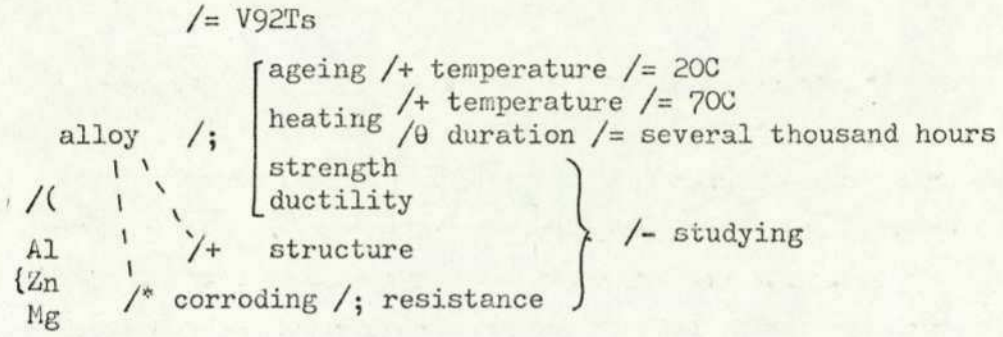
9. 7012 14/0478

The structure, strength and ductility of the alloy aged at 20 C for several months were studied after subsequent heating to 70 C for several thousand hours. The YS and UTS of the material so treated were appreciably higher than those of samples subjected to stepped ageing processes; however, the crack sensitivity and corrosion resistance worsened. The properties of samples aged at 60 C for 24 and 200 C for 1.5 h changed very little on subsequent heating at 70 C.

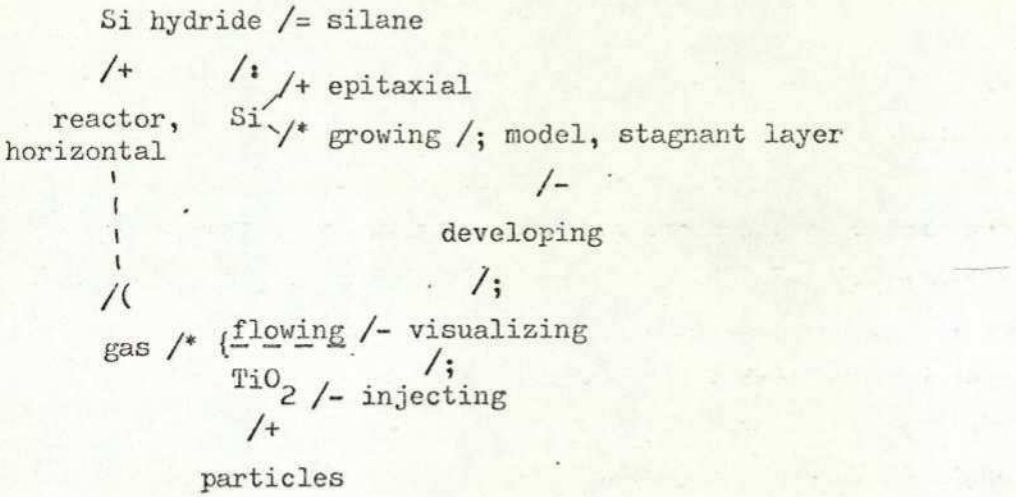
10. 7012 12/1461

Flow patterns were made visible in a horizontal water-cooled epitaxial reactor by injection of TiO_2 particles into the gas flow. From these experiments, a stagnant-layer model was developed with which the epitaxial growth of Si from silane can be described. In the case of a nontilted susceptor, the model predicts an appreciable nonuniformity in thickness along the susceptor, whereas a small angle of tilting of the susceptor should yield a much better uniformity in thickness (2% over a length of 22 cm). Experiments agree well with the theoretical predictions of the model.

9.


/= V92Ts
alloy
/;
ageing /+ temperature /= 20C
heating /+ temperature /= 70C
/θ duration /= several thousand hours
strength
ductility
/+ structure
/* corroding /; resistance
/- studying
Al
{Zn
Mg

10.


Si hydride /= silane
reactor,
horizontal
Si
/+ epitaxial
/* growing /; model, stagnant layer
/-
developing
/;
gas /*
flowing /- visualizing
/;
TiO2 /- injecting
/+
particles

APPENDIX IV

Examples of the browsing techniques

Generic browsing

DOWNWARD BROWSING

Concept in profile	Concept in abstract
magnetic properties	magnetic susceptibility
absorption, optical	absorption coefficient
optical properties	optical constants
hardening	age hardening
working	rolling
deforming	extending
precipitate	carbide
electrochemical properties	dissolving
electrochemical properties	passivity
oscillations	sound waves

UPWARD BROWSING

Concept in profile	Concept in abstract
electrical properties	properties
strength, tensile	strength
recrystallizing	crystallizing
precipitation hardening	hardening
thin film	film
electrical conductivity	electrical properties
stress corrosion	corrosion
nucleating	solidifying
alloy	materials
friction	properties

Conceptual browsing

1. Profile

/: phase /* precipitating

Abstract

/+ solid solution /; supersaturating
/*
decomposing

2. Profile

alloy /- precipitation hardening

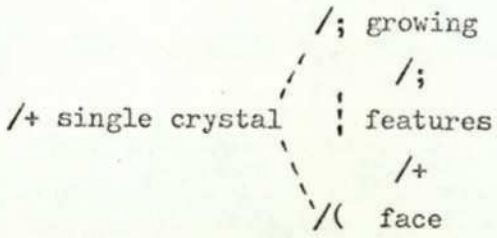
Abstract

/: phase /* precipitating
alloy /;
/* embrittling /- studying

3. Profile

/+ crystal /(structure

Abstract



4. Profile

metal /(electrical conductivity

Abstract

electrical conduction /* metal

5. Profile

alloy /(properties

Abstract

electricity /* transporting /- studying
/+
alloy

6. Profile

Si /; impurity

Abstract

Si /; doping /; P

7. Profile

H
/-
alloy /* absorbing

Abstract

alloy /- embrittling /; H

8. Profile

Al oxide /+ structure

Abstract

Al oxide /+ film /- examining /; electron microscope

9. Profile

Au /; electrochemistry

Abstract

Au /= electrode /; reactions

10. Profile

metal /+ thin film [epitaxial] /* growing

Abstract

Si /- depositing /; kinetics / mechanism

/+

film [epitaxial]

Condensations

Note: Omitted concepts and relations enclosed within square brackets.

Single jumps

Cu [/+ crystal] /; point defects

Ag [/+ structure] /- studying

Si [/(surface] /; defects

alloy [/: phase] /* precipitating

Mo [/(surface] /* heat transfer

alloy /+ [crystal /( ] structure

Al [/: film] /; corroding

alloy [/+ solid solution] /: precipitate

alloy /+ [electrolyte /=] salt

alloy [/+ eutectic] /* solidifying

Double jumps

alloy [/+ film /* growing] /- studying

alloy [/* ageing /; mechanism] /- studying

alloy [/+ solid solution /: precipitate] /* precipitating

Ni [/+ crystal /= whiskers] /* nucleating

Au [/= electrode /(surface] /; roughness

alloy [/: phase (metastable) /* precipitating] /- studying

Ta [/: Ta oxide] /+ [crystal /( ] structure

alloy [/(solid solution /+ metastable] /+ structure

alloy [/(phase (metastable) /: particles] /* precipitating

GaAs [/+ structure] /; [imperfections /( ] dislocations

Triple jumps

Ag [/+ film /(surface /(structure] /- studying

Nd [/+ thin film /+ surface /(work function] /- studying

alloy [/(phase /(SnSb /+ particle] /* coarsening

semiconductor [/* electric transport /; studying /; methods]
/;
review

Quadruple jumps

Zn [/(atom /* ionizing /(spectra /(lines] /; optical
excitation functions

Cu [/+ polycrystal /+ orientation, preferred /* changing /; mechanism]
/;
slip

Jumpbacks

[tapes /( ] /; preparing
V_3 Ga

[metals /=] /(critical temperature
superconductor

[metal /+] thin film
/*
ferromagnetic resonance

[alloy /=] superconductor
/+
[magnetic field /=] critical magnetic field

Jumpbacks (continued)

[metal /+] thin film
/(
magnetic properties

[metals / alloys /+] liquid
/(
surface energy

Cu /- pressure, hydrostatic
/* /;
crystallizing [/; kinetics /* changing]

[metal /+ film /; profile /= Al] /- [studying /;] electron microscope

Related browsing

Concept in profile	Concept in abstract
electron microscope	field emission electron projector
cracks	cracking, micro-
strengthening	hardening
strengthening	strength
crystallizing	crystal
theory	minimum
fracture	strength, ultimate tensile
order	ordering
magnetic properties	domain
electrochemical properties	oxidising [electrochemical]
strength, tensile	strength, yield
specific heat	thermal expansion
critical temperature	superconductor
electrochemical properties	reactions, electrochemical
kinetics	mechanism