A COMPARATIVE EVALUATION OF SEARCHING BY CONTROLLED LANGUAGE AND NATURAL LANGUAGE IN AN EXPERIMENTAL N.A.S.A. DATA BASE

By: C.W.Cleverdon
Cranfield Institute of Technology
Cranfield, Bedford, England

Technical Management: W. A. Martin
E.S.A.Space Documentation Service
ESRIN, Frascati,Italy

July 1977

ESA Contract Number: I/432

EUROPEAN SPACE AGENCY

CONTRACT REPORT

ABSTRACT

An evaluation test was made of an experimental data-base prepared by the Space Documentation Service of the European Space Agency, consisting of some 44,000 items from N A S A STAR for 1973 and 1974. With this data-base it was possible to search on natural language terms in the titles and abstracts, in addition to the normal searches on controlled language index terms. The on-line searches were carried out at four centres, each centre being responsible for ten questions, with two searches in the alternative search modes being made by different people for each question. Up to twenty-five documents retrieved in the two searches for each question were sent to the originator of the question for relevance assessment.

The results are presented in a number of different ways, but in every case the natural language searches showed a significantly higher recall ratio than the controlled language, with little difference in the precision ratios. It is suggested that the main reason for the superiority of natural language searching is the greater exhaustivity of the abstracts as compared to the indexing.

CONTENTS

INTRODUCTION

The Space Documentation Service of the European Space Agency has for many years been responsible for providing a service in Europe with the NASA data bases. In the main this operates as an on-line interactive information retrieval service through various national centres with direct lines to the S.D.S. Computer Centre in Frascati, Italy. Each of these national centres can be accessed by authorised users in their own country.

The NASA data base comprises the contents of both Scientific and Technical Aerospace Reports (STAR) and International Aerospace Abstracts (IAA). It contains the usual bibliographic details plus the indexing entries of some 60,000 research reports, journal articles, conference papers etc. annually, and can be searched on-line in the conventional manner. The NASA system has a number of interesting features. Unlike most of the data bases of comparable size, it is product-oriented rather than subject-oriented. The result, as can be seen by the list of subject categories given in Fig. 1, is that the collection consists of papers over a wide range of subjects, which has some interesting implications for this test. For STAR, consisting as it does mainly of documents issued as research reports, the average length of each item will be some five times that of the journal articles which are found in most systems. While it does not necessarily follow that the information content also varies by a factor of five, yet it must be significantly greater than is normally found, and again this has some interesting implications.

The indexing is done using a controlled language thesaurus. Wall (Ref. 1) reports that, from the commencement of operation in 1962, there was 'an indexing philosophy closely related to that of the Uniterm system' and the indexing was virtually free of any constraints. In 1966 it was decided that a vocabulary control authority was required, and the NASA thesaurus, adopting terminological conventions developed by the Engineers' Joint Council, was used from January 1968. The effect of this was that, from an average of some 18 free terms, the number of index entries was reduced to nine. Wall reports that the effect was to give a significant improvement in the output that was generated for the SDI searches that were then being run.

The publication covering research reports, the bi-monthly STAR, consists of the abstracts of the items arranged according to the subject categories given in Fig. 1, together with author, subject and source indexes. These latter are accumulated half-yearly and annually. In the same way as the National Advisory Committee of Aeronautics fifty years ago, and subsequently NASA, set and maintained a standard, both in presentation and format, for their own reporting which was never matched elsewhere, so with STAR, NASA has for many years past been responsible for a secondary publication which in every way is, with one possible exception, outstandingly good.

The possible exception lies in the indexing. To say this is in no way to denigrate the standard of indexing, but is merely to emphasise that, in the very nature of the process, indexing cannot, in the absolute sense, be perfect. In indexing, one man's meat is another man's poison. In one case not indexing a particular aspect may result in failure to retrieve a relevant document, while in another search, the indexing of the same aspect might result in the retrieval of non-relevant documents. Certainly the standard of NASA indexing compares favourably with that of any other major system, but in some work reported by Cleverdon and Kidd (Ref. 2) examples were found of unsatisfactory indexing in the NASA system which must inevitably degrade the retrieval performance.

01 Aeronautics (General)
02 Aerodynamics
03 Air transportation and safety
04 Aircraft communications and navigation
05 Aircraft design, testing and performance
06 Aircraft instrumentation
07 Aircraft propulsion and power
08 Aircraft stability and control
09 Research and support facilities (air)
12 Astronautics (General)
13 Astrodynamics
14 Ground support systems and facilities (space)
15 Launch vehicles and space vehicles
16 Space transportation
17 Spacecraft communications, command and tracking
18 Spacecraft design, testing and performance
19 Spacecraft instrumentation
20 Spacecraft propulsion and power
23 Chemistry and materials (general)
24 Composite materials
25 Inorganic and physical chemistry
26 Metallic materials
27 Nonmetallic materials
28 Propellants and fuels
31 Engineering (General)
32 Communications
33 Electronics and Electrical engineering
34 Fluid mechanics and heat transfer
35 Instrumentation and photography
36 Lasers and masers
37 Mechanical engineering
38 Quality assurance and reliability
39 Structural mechanics
42 Geosciences (general)
43 Earth Resources
44 Energy production and conversion
45 Environment pollution
46 Geophysics
47 Meteorology and climatology
48 Oceanography
51 Life sciences (General)
52 Aerospace medicine
53 Behavioral Sciences
54 Man/system technology and life support
55 Planetary biology
59 Mathematical and computer sciences (general)
60 Computer operations and hardware
61 Computer programming and software
62 Computer systems
63 Cybernetics
64 Numerical analysis
65 Statistics and probability
66 Systems analysis
67 Theoretical mathematics
70 Physics (general)
71 Acoustics
72 Atomic and molecular physics
73 Nuclear and high-energy physics
74 Optics
75 Plasma physics
76 Solid-state physics
77 Thermodynamics and statistical physics
80 Social sciences (general)
81 Administration and management
82 Documentation and information science
83 Economics and cost analysis
84 Law and political science
85 Urban technology and transportation

Fig 1. cont'd

88 Space sciences (general)

89 Astronomy

90 Astrophysics

91 Lunar and Planetary Exploration

92 Solar physics

93 Space radiation

99 General

Fig. 1 List of subject divisions of NASA STAR

A considerable amount of experimental data now exists which shows that, as stated earlier, indexing can never be perfect, however strict the vocabulary control or management control may be. Such experimental work has also shown that the use of natural language with free text searching on the title and abstract can be as effective, and sometimes more so, than searching on controlled language terms assigned by an indexer. Such evidence has come mainly in experimental tests, and there are many reasons why in an operational system it may neither be possible to replicate the results, nor might it be practical to use such a method for real searches. On the other hand the possibility of using natural language free text has many advantages for the users. Not only can search terms be obtained more easily but it is possible to switch from one data base to another without having to generate a new set of search terms.

The background to this test is that the possible advantages of natural language free-text searching in a multi-file on-line data base were sufficiently attractive to S.D.S. that they decided to initiate an investigation which would show whether natural language searching was 'practical' with the NASA data base when compared with the controlled language searching as a yardstick. They also wished to test whether a machine prepared listing of Associated Concepts would significantly assist natural language searching. The term Associated Concepts' is intended to describe a search aid for use with natural language search formulation which would be analogous with the familiar 'related terms' feature of a conventional thesaurus.

SUMMARY OF TEST DESIGN (see Appendix A)

A test collection of some 44,000 items covering the input to NASA STAR for 1973 and 1974 was prepared such that on-line searching could be done using either controlled language indexing or natural language on titles and on abstracts, or on any combination of these. In addition, to assist in the formulation of natural language searches, an off-line aid was available in the form of a computer print-out. This was the Associated Concepts (A.C.) file, which had been prepared from the title words of the source data.

It was intended that the test searches should be carried out by the information staff of six centre in direct contact with S.D.S. at Frascati, namely

Royal Institute of Technology, Stockholm, Sweden
Lund University, Lund, Sweden
ESTEC, Noordwijk, The Netherlands
Technical Library of Denmark
Z.L.D.I., Munchen, Germany
Technology Reports Centre, U.K.

At each centre the staff would be responsible for obtaining ten search questions from scientific or technical staff within the centre, with the exception of the Technology Reports Centre in England, where the question would come from staff in user organisations.

It was required that for each question there should be two searches made by different people. The pattern of searches would reflect the four modes in which they were to be carried out, namely with Controlled Language index terms, Natural Language searching on words in the title and abstract, Controlled Language and Natural Language combined and, finally, Natural Language also making use of the Associated Concepts File. The pattern of searches would be the same at each centre, being as follows:

QUESTION	SEARCHER A	SEARCHER B
01	Controlled Language and Natural Language	Natural Language and Associated Concepts File
02	Controlled Language	Natural Language
03	Controlled Language and Natural Language	Natural Language and Associated Concepts File
04	Controlled Language	Natural Language
05	Controlled Language and Natural Language	Natural Language and Associated Concepts File
06	Natural language	Controlled Language
07	Natural Language and Associated Concepts File	Controlled Language and Natural Language
08	Natural Language	Controlled Language
09	Natural Language and Associated Concepts File	Controlled Language and Natural Language
10	Natural Language	Controlled Language

The output for each search would be sent to the Project Director, who would be responsible for sending, to the originator of the question, the output or a selection

of 25 items if the total output exceeded this number. The questioner would be asked to decide the relevance of the documents; from these decisions comparative perf ormance figures would be obtained for the various search modes, with failure analysis made to determine the reasons for the comparative results and also whether any recommendations could be made to improve performance.

The test design was influenced by three factors. First, the wish of S.D.S. that it should be completed at an early date, secondly the load which could be put on the co-operating centres, and finally the view of S.D.S. that they did not require the most precise results. This last point was because it was considered that the over-riding advantages of the use of natural language as a common retrieval language to enable easier searching across all the several files which comprise the RECON database were such as to predispose S.D.S. towards its use unless it was shown to be impractical. Thus a five per cent, or even possibly a ten per cent difference in performance in favour of controlled language would not necessarily influence S.D.S. against natural language since, in overall terms i.e. in relation to the total database, the potential rewards appeared to be very promising. There was, of course, no intention to discontinue the provision of controlled language indexes but these would remain, as now, unique to each specific file.

For these reasons, it was agreed that the test which, in other circumstances, might have been designed to have at least 200 searches, could be carried out satisfactorily with the smaller number planned.

TEST OPERATION

Following the acceptance of the final proposal a meeting was held at ESTEC with representation from each of the co-operating centres and agreement was reached on the methods to be used in carrying out the test. Following this, the Project Director sent a letter to the centres indicating the detailed procedures to be followed, together with a supply of forms to be used locally. The set of instructions is shown in Fig. 2. Form A, (a completed copy of which is shown in Fig 3), was to indicate the search question and the name of the questioner.

A separate copy of Form B (Fig. 4) was provided for each search of each question and indicated who had carried out the search and the language used. For the later detailed analysis, it was necessary for a full record to be kept of the actions carried out in each search, showing for example what terms had been used and which documents had been consulted etc. Therefore, on Form B the searchers were asked to add comments of their own regarding the reasons for the actions which they had taken in the course of the search.

The questions (see Fig. 5) were duly obtained at the centres and the searches carried out by two persons according to the instructions. Each search finished with a request that the preferred search set should be printed in modes 1 and 4, namely document reference numbers and the complete printout of title, abstract, bibliographic data and index terms. It was arranged that the computer printout should be sent direct to the Project Director.

When the completed copies of Form A were received at Cranfield, a letter was sent to the questioner explaining what would later be required in regard to assessing the search output, and also asking him to list on the form provided any papers of which he was already aware as being relevant to his question. An example of a completed form is shown in Fig. 6.

Additionally a manual search was made by the Project Director at Cranfield. This was normally done using STAR for 1973 and 1974, either by consulting the index or sometimes by looking in the appropriate section of the abstracts. It was accepted that, in doing the former, there was the danger that the results might show a bias in favour of the controlled term searching since the manual search would of necessity, be using index terms which would also, presumably, be used in the Controlled Language search.

On receipt of the computer printout, the abstracts of the documents to be assessed were sent to the questioner. This set of abstracts consisted of three groups, these being documents (up to a maximum of 5) retrieved by a manual search, documents retrieved by the Controlled Language search and documents retrieved by the Natural Language search, but if in total these did not exceed 25 in number, all would be sent. If they exceeded 25, then a random sample would be selected of the documents retrieved by the two computer searches. Normally this would be pro-rata for the number of documents retrieved by each system, but if there were a large discrepancy (for example 10 by Natural Language and 100 by Controlled Language) then a minimum of six were selected from the minor set.

It is important to appreciate that, where a selection of documents was required, it was made on a random rather than a selected basis. This was so that it would be possible to calculate the retrieval performance of the whole set of retrieved documents on the basis of the assessed sample. The ability to do this was necessary for the preparation of the final figures. In doing this it was accepted that in an individual search, where say, 60 items were retrieved and of the 12 documents assessed, 75% were found to be relevant, there would be a relatively wide margin of error in asserting that 75% of the

EVALUATION TEST OF ALTERNATIVE SEARCH TECHNIQUES IN NASA DATA BASE

TEST PROCEDURES

1. The first stage is to obtain the co-operation of ten persons who have or will submit search questions and who agree to carry out the subsequent relevance analysis. For this purpose, Form A should be used. The completed copies of these forms should be sent to the Project Director as soon as possible, after having entered the question number (see 2A below).

2. The question having been received, the search can be carried out as soon as ESRIN advise that the test file is ready. The following is a procedure to be followed for each question, based on the expectation that there will be ten questions and two persons to do the searching at each centre.

   a) Twenty copies of Form B are enclosed, two to be used for each question. The forms indicate the particular search strategy to be adopted and are serially coded for each question, the code representing the centre and the individual search. Enter this search code on Form A, so that the question and search can be matched.

   b) Enter on Form B the name of the search operator carrying out each particular search.

   c) The two search operators should agree on the exact requirements of the question (discussing this with the originator of the question if this is necessary and practical). They should also agree on the minimum requirements of the question. As an example of this, assume a question asks for information on "a method of calculating the aerodynamic load distribution on swept-back wings with fuselage". To obtain 100% recall on such a question might involve searching on all types of wings or on "load distribution", but this would probably retrieve an unacceptably large mass of non-relevant documents and a more restrictive search would be "calculation" and "load distribution", or alternatively "load distribution" and "wings". I am not concerned with suggesting exactly what the minimum should be for such a hypothetical question, but it is essential that there should be reasonable agreement between the two search operators as to the acceptable level for each question.

   d) While the main purpose of the test is to evaluate any differences in performance between controlled languag and natural language, there are two additional requirements. The first of these is that we should test the possible effect, when searching by natural language, of using the list of associated concepts, which it is intended will be supplied to each centre by ESRIN. The second point relates to the effect using both controlled and natural language in a single search. Therefore, the pattern of searches should be as follows:

Question	Searcher A	Searcher B
1	C.L. and N.L.	N.L. with A.C. File
2	C.L. only	N.L. only
3	C.L. and N.L.	N.L. with A.C. File
4	C.L. only	N.L. only
5	C.L. and N.L.	N.L. with A.C. File
6	N.L. only	C.L. only
7	N.L. with A.C. File	C.L. and N.L.
8	N.L. only	C.L. only
9	N.L. with A.C. File	C.L. and N.L.
10	N.L. only	C.L. only

Fig. 2 Test Instruction Sheet for operations at Centres

The obvious problem in relation to the two additional requirements is whether they have to be treated as obligatory or optional. Clearly, where the instruction is to use only natural language or controlled language (as with questions 2, 4, 6, 8 and 10) then the instruction should be rigidly adhered to. However, this would not always be satisfactory in the other two situations. An operator may find, for example, that a perfectly satisfactory search can be done using only controlled language terms, and has no need to use natural language terms. In this case, I would suggest that the instruction be treated as optional; that is to say the search would be conceived in terms of a controlled language search, but the freedom also to use natural language should be taken whenever it appears that it would improve the search. In the case of the natural language searches using the concept file, again I would suggest that it must be considered as an optional accessory, and that the associated concept file should be consulted in all reasonable cases.

e) Carry out the search in the usual way except that if a visual display unit is being used, record on the printed search record information regarding any documents which are displayed but which will not be included in the final printout. Also remember to ask that the printout by ESRIN should be in both format 1 and format 4.

f) Enter on Form B the time taken by the searcher in any pre-terminal activities associated with the search but not including the joint discussion on the definition and scope of the question. Enter also on Form B the time spent at the terminal for the search.

g) Enter on Form B any notes which might be helpful for changes in the search strategy, cross referencing these with the appropriate part of the printout. Finally any general comments on matters relating to the search language that might be of interest.

h) Send Forms B and the appropriate printed search record to Project Director.

3. Copies of the printout for each search will be sent by ESRIN direct to the Project Director.

4. As soon as the copies of Form A have been received by the Project Director, he will write to the individual concerned and ask him to list any known relevant documents. Additionally a manual search will be made at Cranfield to try to locate a few of the documents which appear to be relevant, and which can be used for determination of the recall ratio.

5. If the printout for a given question contains more than 30 citations, a random subset of 30 will be obtained and the abstracts of these will be sent to the questioner for relevance assessment. If there are less than 30 references, all will be sent to the originator of the question. Information on this point will also be sent to the appropriate Centre Director.

6. The relevance assessments having been returned, the performance figures will be calculated for each search and failure analysis for a selection of the failure documents (that is the relevant documents that were not retrieved, or the non-relevant documents which were retrieved). This will be related to the search strategy and a final analysis made for each question.

7. In so far as time permits, copies of the analysis in (6) will be sent to the appropriate Centre Director for comment before the final report is written.

8. Copies of the final report will be made available to Centre Directors prior to proposed meeting at Frascati in April.

FORM A

EVALUATION TEST OF ALTERNATIVE SEARCH TECHNIQUES IN NASA DATA BASE

TEST PROCEDURES

QUESTION ORIGINATOR

Name ....Mr.J..Walker.....

Adress ESTEC, Domeinweg,
.....
NOORDWIJK, Holland.
.....
.....

SEARCH QUESTION

Methods and apparatus for testing rocket fairing separation
.....
under low pressure in vacuum.
.....
.....
.....

SEARCH NUMBER
(taken from Form B)

Qu. 7.

Fig. 3 Completed copy of question form

FORM B

EVALUATION TEST OF ALTERNATIVE SEARCH TECHNIQUES IN NASA DATA BASE

TEST PROCEDURES

Search number: 1-07 C.L. and N.L.

Search operator: B

Date of search: 13/1/77

Pre-terminal search time: NIL

Search time at terminal: 50 mins.

Notes on search strategy (Relate the numbers to the appropriate section of the print record. Use extra sheet if necessary)

1. Expand * Selection of terms * Combining concepts low precision * tests
2. Grouping "faining" terms
3. Grouping "jettison" + test terms
4. Displayed set 28 for more relevant terms
5. Set 31 produced nothing interesting
6. Neither did set 33
7. Finally printed set 34 "faining" terms + "jettison" terms as containing "tests" implicitly
8.

General Comments:

Fig. 4 Completed search strategy form

Centre: ESTEC, Holland

1-01 Types of corrosion (including galvanic coupling) normally to be encountered in a gold-nickel alloy joint brazed on 18/8 chromium nickel steel (stainless) tubing carrying water or freon for cooling.

1-02 The use of fibre optics in all forms of communication including telecommunications. Coherent laser optics in general can be excluded.

1-03 Software for fault-tolerant computing. The writing of programs to detect and correct computer errors during operation and to minimise fault-propogation and avoid breakdown.

1-04 Communications between satellites and mobile stations: ships aircraft and all mobile stations including other spacecraft.

1-05 Thermal control coatings of low electrical conductivity for use on spacecraft. All types of coating can be of interest and this aspect must be explored fully. Requestor wishes to have information on the electrical properties of all types of coating in order to evaluate those which would be most suitable to avoid the build-up of electrical charge on the satellite. Reader would prefer a largish noisy print in order to have a large choice to evaluate himself.

1-06 Spacecraft contamination caused by use of electric propulsion and how to measure it. This includes electrostatic, ion and solar-electric propulsion, caesium engines, mercury and ion thrusters. Contamination can be thermal or optical, can cause degradation of the solar array, short circuits, chemical reaction between mercury and caesium, or environmental contamination of the earth.

1-07 Methods and apparatus for testing rocket fairing separation under low pressure in vacuum.

1-08 Study of low frequency or decametric radiation from Jupiter and especially the modulation lanes resulting from it. Decimetric radiation is not relevant.

1-09 Scientific objectives of the Large Space Telescope (now called the Space Telescope). The missions and experiments and also a description of the instruments to be used are both required, as the limitations of the instrumentation available can dictate the experiments proposed. Suggestions include cosmological studies faint objects spectrograph, infrared photometer, planetary camera, astrometry experiments. Reader is not interested in other proposals for astronomical research where the work would be done via other projects.

1-10 Vibration testing techniques for space structures using modal analysis methods. This is a method to determine the vibration modes of complex structural systems.

Fig. 5 List of search questions by Centres

Centre: Technology Reports Centre

2-01 Friction drive transmission systems

2-02 Aluminium chlorides (Tri-chlorides and mono chlorides) - Thermodynamic properties and kinetics including entropy, free energy, specific heat etc.

2-03 CO_2 in the atmosphere

2-04 Control of hydrofoil craft in a random seaway to improve passenger side comfort

2-05 Test data on low reynolds number wing aerodynamics

2-06 Errors in gyroscopic flight instruments.

2-07 Use of microprocessors or programmable calculators in practical applications (e.g. process control)

2-08 Cometary dust and comet structure particularly comet tails

2-09 Automatic control of satellites by computer

2-10 Aerodynamic noise produced by subsonic flow of air or gas through a jet or nozzle (not environmental noise)

Centre: University of Lund

4-01 Methods of range determination to a moving target.

4-02 Aerosol induced changes in photobiologically essential spectral components of visible light.

4-03 Studies of micro- and local- (meso-) climatic effects by remote sensing.

4-04 Rock density measurements and prospecting by means of underground and sea level measurements of cosmic ray intensity.

4-05 All aspects of positive pumps; lobular and gear pumps in particular.

4-06 Routing and flowcontrol in computer communication networks.

4-07 Isotopic composition of cosmic ray nuclei. Measurements and theories.

4-08 Atomic Ba, Ca, Sr, Mg. Energy levels (Stark effects, Seman effects) and photoionization.

4-09 Studies of cultivated crops in United States by use of remote sensing.

4-10 Decompression sickness. Factors influencing and individual variations.

Fig. 5 List of search questions by Centres (cont)

Centre: Royal Institute of Technology
STOCKHOLM

5-01 Diffusion of free carriers with quadratic and/or cubic recombination in semiconductor devices; diffusion of free carriers at high carrier concentration.

5-02 Inertial navigation

5-03 Electromagnetic and acoustic detection of buried objects (with special application to avalanche victims)

5-04 Methanol-air fuel cells

5-05 Performance of solar cells in a terrestrial environment

5-06 Application and analysis of periodically switched linear circuits.

5-07 Adaptive antenna systems for suppression or cancellation of antenna sidelobe jamming and interference. Properties of criterial and algorithms for adaptive antenna systems

5-08 Synchronization of digital networks.

5-09 Eigenvalues of matrices (especially sparse matrices), least-square algorithms

5-10 Wake vortex, vortex trail, vortex turbulence.

Fig. 5 List of search questions by Centres (cont)

MR J Walker

ESTEC

Search question:

Methods and apparatus for testing rocket fairing separation under low pressure in vacuum.

The following references are known to be relevant to the above question.

1. CENTAUR AC-4 NOSE FAIRING JETTISON TESTS
NASA TM X-52154

2.

3.

4.

Please return this form to: Mrs B Oldroyd
Library
ESTEC

Fig. 6 Completed known-relevant documents form

total output of 60 were relevant. In fact, in such a case, the standard error would be approximately $\pm$ 16% but when such figures are calculated over the whole range of searches in this test, the standard error will only be $\pm$ 3%.

Copies were made of the abstracts to be assessed; each was given a document number and, together with a letter and an assessment form were sent to the questioner. The request was that the documents should be assessed for their relevance to the question as submitted. The questioner was asked to indicate on the form whether he considered the document to be of relevance 1, relevance 2 or non-relevant, these being intended to show:

1. a highly relevant document which the user would definitely wish to see;

2. an item which is related to the search query but is not likely to be so useful as the items in category 1.

When it was decided that an item was not relevant, it was requested that the questioner should, in the space provided, add a brief comment on why he made this decision. A completed copy of an assessment form is shown in Fig. 7.

It was requested that this form should be returned either direct to the Project Director or through the local organiser.

QUESTION 1-07 — 17 Mr J Walker

Methods and apparatus for testing rocket fairing separation under low pressure in vacuum

Document	Relevance 1	Relevance 2	Non-relevant	Comments
1			✓	SEPARATION (AERODYNAMIC) ≠ SEPARATION (MECHANICAL
2	✓			
3	✓			
4	✓			
5			✓	AS (1) ABOVE
6	✓			
7		✓		
8			✓	AS (1) ABOVE.
9		✓		
10			✓	AS (1) ABOVE
11			✓	" " ?
12			✓	" " ?
13	✓			
14			✓	AS (1) ABOVE
15			✓	" "
16			✓	" "
17			✓	" "
18				
19	✓			
20		✓		
21			✓	AS (1) ABOVE.
22		✓	✓	THE REFERENCES IN THIS DOCUMENT MAY LEAD TO THE TEST PROGRAMME FOR THIS TYPE OF ROCKET
23			✓	AS (1) ABOVE.
24		✓		
25				

Fig. 7 Completed relevance assessment form

METHODS OF PRESENTATION OF RESULTS

For each question, a listing was made of all the documents retrieved in the two search modes, as well as the references found in a manual search (see example in Fig. 8). The complete set of such sheets is included as Appendix B.

An example of a complete Master Record Sheet is shown in Fig. 9, relating to search 1-07. In the first column of this sheet are the file numbers of the documents sent to the questioner for relevance assessment. The letters in the second column relate to the search mode for which the particular document was being assessed. In Fig. 9, for example, most documents were included either for Controlled Language alone or for Natural Language alone, but it will be noted that documents 420544, 410242 and 329962 were assessed both in regard to Controlled Language and Natural Language, and that documents 416978, 422510 and 314853 were assessed against the manual search in addition to Controlled Language or Natural Language.

When the completed user assessment sheet (see Fig. 7) had been returned, the relevance decisions were transferred to the master record. Next a check was made against the list of retrieved documents (Fig. 8) to ascertain which documents that had been assessed as relevant had been retrieved by another system than that for which they were assessed. With search 1-07 it can be seen, by the entry in the final column, that (3) 416978 was also retrieved by Natural Language and that (13) 422510, (18) 410242, (19) 333850 and (24) 314853 had also been retrieved by Controlled Language. A check of the printout of the abstracts was then made to ascertain whether the documents retrieved by Natural Language searches would have been retrieved if the search had been restricted to titles, instead of both titles and abstracts. This required consulting the computer search records; in this particular case it was decided (very unusually) that seven of the eight relevant Natural Language documents would have been retrieved on titles. This is indicated by the letter T in the final column.

The data shown in the Analysis section of Fig. 9 were obtained from the computer printout and from the completed records of the various searches.

The performance figures were obtained from the results recorded elsewhere in the record sheet. These were calculated for relevance 1 and for relevance 1 and 2 documents and are presented in the usual measures of recall and precision ratios.

In an operational evaluation, there is no practical way in which a correct recall ratio can be obtained, for it is clearly impossible for questioners to look through 44,000 documents and make a decision as to whether each item is or is not relevant to his request. Since the true recall cannot be obtained, some alternative procedure has to be used, and the two qualified recall measures used in this test are known as 'base recall' and 'matched recall'. The former measure was devised for the test of the MEDLARS system (Ref. 3) and proved entirely satisfactory. It requires that relevant papers should be found outside of the system being tested. The base recall ratio is the percentage of such documents retrieved by the test system. In question I-07 it can be seen that there were two relevance 1 documents and one relevance 2 document retrieved by the manual search, All three of these documents were retrieved both by Controlled Language and Natural Language searches so in all cases for this question the base recall ratio was 100%.

The matched recall ratio is a variation of the base recall ratio and can be used when,

Question 1.07

Controlled language

434458 422510 420544 416978 410242 333850 332782 331785
329962 324943 324676 320937 318043 314853 311897

Natural Language

434337 433307 422639 422510 422508 422503 421898 421522
420544 416978 410825 410242 333850 331785 331705 330250
329962 326992 318846 318043 313288 313021 312492 314853
311383

Manual Search

416978 422510 314853

Question 1.08

Controlled Language

435263 413546 431288 314440

Natiral Language

435263 431288 430275 427350 426275 422488 421421 415487
413546 411628 411626 411624 332659 327746 320847 314833
314440 310816

Manual Search

411626 413546 421421 435263 314440 430275 314833 332659

Fig. 8 Output for Questions 1.07 and 1.08

QUESTION 1-07

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	434458	C			✓	
2	420544	CN	✓			T
3	416978	CM	✓			N
4	410242	CN	✓			T
5	332782	C			✓	
6	329962	CN	✓			T
7	324676	C		✓		
8	318403	C			✓	
9	311897	C		✓		
10	331755	C			✓	
11	434337	N			✓	
12	422639	N			✓	
13	422510	NM	✓			C T
14	422508	N			✓	
15	421898	N			✓	
16	421522	N			✓	
17	410825	N			✓	
18						
19	333850	N	✓			C T
20	331705	N		✓		
21	330250	N			✓	
22	318846	N		✓		
23	313288	N			✓	
24	314853	NM		✓		C T
25						

ANALYSIS

	C'L.	N.L.
Items Retrieved	15	25
Overlap	9	
Search time	34	37
Pre-search time		45
Total sets	34	36
Search sets	25	25
Combine sets	9	11

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	4	5	5
Relevance 2	2	3	1
Non-relevant	4	8	0
Ratios			
Relevance 1 & 2			
Precision	60	53	
Base Recall	100	100	
Matched Recall	82	82	
Relevance 1			
Precision	40	31	
Base Recall	100	100	
Matched Recall	100	100	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
416978	314853
422510	

SEARCHER MODE

A NL + Coverdt

B CL + NL

Fig. 9 Completed Master Record Sheet for Question 1-07

as in this test, a comparison is being made between two systems. It assumes that the sum of relevant documents retrieved by either of the two test systems represents all relevance documents in the collection. The matched recall ratio is the proportion of relevant documents retrieved by one system as against the total retrieved by both systems. The use of this measure in the present test was complicated by the fact that relevance decisions were only available for a sample of the retrieved documents since, as discussed earlier, no more than 25 references were ever sent to a questioner.

There are three possible ways of determining the matched recall ratio. To understand these, it is necessary to consider the limitations of the relevance decisions in regard to those searches where the total combined output was more than 25 different references. As considered earlier, in such a case the output sent to the questioner was selected at random from each output, maintaining a proportion for the individual outputs, as for example, in Fig. 8, where the references selected for relevance judgements for a given question are underlined. This indicates, for example, that, amongst others, 420544 was selected for both Controlled Language and Natural Language, but that 333850 was only selected for Natural Language, although it was also retrieved by Controlled Language. The documents judged relevant (see Fig. 9) and the system in respect of which the relevance judgement was made were as follows:

420544	CL and NL
416978	CL
410242	CL and NL
329962	CL and NL
324676	CL
311897	CL
422510	NL
333850	NL
331705	NL
318846	NL
314853	NL

This shows that 11 documents were judged relevant, of which 3 were assessed as being Controlled Language alone, 5 as being Natural Language alone, and 3 as being both Natural Language and Controlled Language; the remaining 12 were, of course, not relevant.

The three possible methods of determining matched recall are as follows:

Method 1

This considers the results solely from the relevance judgements as determined for each method, thus for Controlled Language, the matched recall ratio would be 6/11 = 54%, and for Natural Language it would be 8/11 = 72%.

Method 2

This takes account of all retrieved documents known to be relevant. By reference to Fig. 8, it can be seen that reference 416978 was also retrieved by Natural Language, and 422510, 333850 and 314853 were retrieved by Controlled Language. Thus each system retrieved 9 of the 11 known relevant documents, so the matched recall ratio would be 82% for each system.

Method 3

As previously argued, the estimated total number of relevant documents retrieved by each system can be obtained by extrapolating from the proportion assessed and judged relevant. This third method determines the matched recall ratio on the basis of such figures. In our example for search 1-07, 15 documents in all were retrieved by Controlled Language and 25 by Natural Language, so it is assumed that the total relevant retrieved by CL is (6/10) x 15 = 9, and for Natural Language it would be (8/16) x 25 = 12.5, giving a combined total of 21.5.

However, since it is known that seven relevant references were retrieved by both systems, these should be deducted from the above figure which now becomes 14.5. On this basis, the comparative matched recall ratios would become 62% for Controlled Language and 86% for Natural Language.

Compared to Method 2, this final method lowers the recall for the system retrieving fewer documents and increases it for the system retrieving more documents. Even in this case the two methods show a significant difference; depending on variations in the number of documents retrieved, the difference could be quite large. In one search Natural Language retrieved 2 items, and Controlled Language retrieved 150. Both Natural Language documents were judged relevant; of the 23 assessed for Controlled Language, 3 were judged relevant. Controlled Language also retrieved the two Natural Language assessed documents. The matched recall ratios would be

	C.L.	N.L.
Method 1	45%	60%
Method 2	40%	100%
Method 3	10%	100%

It is obvious that matched recall ratio is not an entirely satisfactory measure to use in these circumstances. Of the three methods outlined, the first is rejected on the grounds that it both gives undue advantage to the system retrieving relatively few references and penalises the system retrieving more references. Method 3 is undoubtedly the fairest, for, correctly calculated, it would represent the situation if all the output had been assessed for relevance. The problem lies in its correct calculation, particularly when there is a combination of a large retrieval of documents, with a high overlap between the two search outputs.

Method 2 has the weakness that it is advantageous to the system retrieving fewer items, but, as can be seen from the examples given above, it does not handicap the system retrieving more references. It was known that, on average, the Controlled Language searches were retrieving fewer items than the Natural Language searches. In view of the objectives of the test, it was felt that any measure used should be one which would show a bias towards, rather than against, Controlled Language. Method 2 has such a bias and for this reason, and also because of the arguments that could develop in regard to the application of Method 3, Method 2 was selected for the Matched Recall Ratio.

In regard to question 1-07, six documents in all were assessed as relevance 1, and five documents assessed as relevance 2. Both Controlled Language and Natural Language retrieved all relevance 1 documents, so the matched recall ratio was 100% in each case. However, Controlled Language did not retrieve two of the relevance 2 documents, and Natural Language failed to retrieve two relevance 2 documents. This results in a matched recall ratio for 1 and 2 relevance documents of 82% for Controlled Language and 82%for Natural Language.

Recall ratios on their own have little meaning and have to be considered in relation to the precision ratio. There are two methods of calculating precision ratios in an evaluation test, these usually being known as 'average of ratios' or 'average of numbers',

(for a fuller consideration of this matter see Ref.4).* The average of ratios method involves first calculating the precision ratio for each question and then taking an average of the sum of the ratios. The average of numbers involves summing the relevant and non-relevant documents retrieved in all the searches and deriving the precision ratio from these totals.

In theory these two methods can present significantly different results, as can be seen from a hypothetical example of a set of five searches.

	Relevant	Non-Relevant	Precision ratio
(1)	4	1	80%
(2)	3	3	50%
(3)	4	2	67%
(4)	8	2	80%
(5)	2	30	6%

With a total of 21 relevant and 38 non-relevant items, the average of numbers for these five searches gives a precision ratio of 35%, whereas the average of ratios would give a figure of 57%. Experience of a number of tests has shown that the average of ratios usually gives a slightly higher figure, but to my knowledge there has been no case where if System A has a superior precision ratio to System B by one method of calculation, the situation is reversed by the other method of calculation.

On the Master Record Sheets (all of which are included in Appendix C) the precision ratio refers to the individual search. In the presentation of the test results, the precision ratio is given by both methods.

All the percentage figures given in this report are only accurate within the standard error, which is mainly dependent on the number of relevance decisions on which the calculations are based. The standard error can be obtained by the equation

$$o = \sqrt{\frac{PQ}{N}}$$ where P = probability of event, Q = 100-P and N = size of sample.

As an example, assume a precision ratio of 70% obtained on the basis of 100 relevance decisions. In such a case the standard error would be $\sqrt{\frac{70 \times 30}{100}}$ = 4.6%. If it had been based on 500 relevance decisions, the standard error would be reduced to 2.0%. This means that for the individual questions (e.g. Fig. 9) the performance ratios can only be considered accurate at best within ± 16%. However, summing a set of results within a single Centre or by a particular search mode reduces the standard error considerably, and with most sets of performance ratios presented in this report it lies within the range of ± 3% to ± 6%. The approximate standard error is shown for the major presentations of results in this report.

Two of the Centres were unable to complete the test searches, so the results are based on searches carried out at the following Centres.

1. ESTEC
2. Technology Reports Centre
4. Lund University
5. Royal Institute of Technology.

* It is also possible to calculate the recall ratio by either method, but it does not appear to be meaningful to use the 'average of numbers' method when, as in this report, the measure is 'matched recall ratio'. The recall ratio has therefore been calculated by 'average of ratios'

TEST RESULTS

There are many possible ways of deriving results from the data available. Not all possible combinations have been dealt with; for example, it does not seem necessary in this report to consider whether, within a given centre, searcher A had a superior performance to searcher B or vice versa. However, all the data necessary for such further analysis have been included in the report or in the Appendices.

The main performance data from the individual Master Record Sheets has been transferred to the Centre Record Sheets (see Figs. 10a - 10e). These have been grouped according to the four methods of search, namely:

a. Natural Language
b. Natural Language and Associated Concepts File
c. Controlled Language
d. Controlled Language and Natural Language

From the Centre Record Sheets the results have been summarised, again according to the four search modes, and are included as Figures 11a - 11d This data has been transferred to the Mode Record Sheets in Figures 12a - 12d. From these the complete performance figures have been compiled and are presented in Table I.

	PRECISION				RECALL			
	Average of ratios		Average of numbers		Matched		Base	
	Rel.1 %	Rel.1+2 %	Rel.1 %	Rel.1+2 %	Rel.1 %	Rel.1+2 %	Rel.1 %	Rel.1+2 %
Natural Language	31	63	26	54	75	78	64	72
Natural Language and Associated Concepts File	23	52	24	51	76	73	75	79
Controlled Language	40	74	37	70	68	56	49	48
Controlled Language and Natural Language	22	45	19	47	64	71	74	79

TABLE 1. Overall performance figures for search modes. (S. E. $\pm$ 3% to $\pm$6%)

It can be seen that there are only minor variations for the Precision Ratio when calculated by the Average of Ratios or the Average of Numbers. For this reason, future presentation of results will be based on the Precision Ratio being calculated by the Average of Ratios.

In regard to the two methods of calculating the recall ratio it will be noted that there are some major variations particularly in regard to Relevance 1 documents, as for example with Controlled Language (68% and 49%). This is not suprising in view of the fact that the Base Recall Ratio in this case was based on only 37 relevant documents; In this respect, there were some eighty relevant documents for determining the Relevance 1 matched recall ratio for each mode, as against

approximately 180 for Relevance 1 and 2. This means that the latter will give a more reliable figure, and therefore, in the later discussion of results, the Relevance 1 and 2 figures for the matched recall ratio will be used.

A simplification of Table 1 based on these decisions is presented as Table 2.

	Recall	Precision
N.L.	78%	63%
N.L. and A.C. File	73%	52%
C.L.	56%	74%
C.L. and N.L.	71%	45%

Table 2. Performance figures for Relevance 1 and 2 documents (s.e. $\pm$ 3%)

The usual phenomenon of the inverse relationship of recall and precision appears, with Controlled Language having the lowest recall and highest precision ratio. That it does not always come into effect is shown by the figures for Natural Language with the highest recall and also superior precision to the two other other modes, but the difference in recall is hardly significant.

The inverse relationship of recall and precision is a well established phenomenon. Formally stated it says that:

> "Within a single system, assuming that a sequence of subsearches for a particular question is made in the logical order of expected decreasing precision, and the requirements are those stated in the question, there is an inverse relationship between recall and precision, if the results of a number of different searches are averaged" (Ref. 5)

One would not necessarily expect this rule to apply to every question in the present test, since three of the qualifications are not being met. In the strict sense of the word, the searches are not being done 'within a single system'; with two searchers there is no question of a 'logical order of decreasing precision' and it is accepted that a single search can show a deviation from the norm. However, by whichever way the final results are presented, this inverse relationship does appear, and for the later discussion of the results it is important to check that it is valid and has not been influenced by a few abnormal searches.

For this reason an analysis was made of the results to ascertain in which questions the inverse relationship operated as against those questions where one mode showed a superiority for both recall and precision. Based on the figures for Precision and Matched Recall for Relevance 1 and 2 documents, in 27 of the searches the inverse relationship applied, with Controlled Language showing both higher recall and precision in seven searches and Natural Language in five searches.

	Titles			Full Search		
	Rel.1	Rel.2	N.R.	Rel.1	Rel.2	N.R.
Centre 1	18	10	12	25	37	74
Centre 2	11	7	1	29	48	59
Centre 4	21	16	4	37	36	59
Centre 5	17	14	1	30	29	50

Table 3. Document Retrieval by titles as compared to full search

There are two other situations which show the effect of the inverse relationship. The first of these relates to the effect of limiting the Natural Language search to the titles as against searching on the titles and abstracts. In Table 3 are shown the returns for each Centre for documents retrieved in this way with the figures for the full search on abstracts as a comparison, while Table 4 gives the performance ratio for these searches.

	Titles	Full Search
Recall	31%	75%
Precision	87%	53%

Table 4. Performance figures for titles and full search on abstracts with Natural Language (s.e. $\pm$ 4%)

The second aspect relates to the manual searches carried out on the printed indexes. There was no attempt to make a comprehensive search, the intention being merely to find a few possibly relevant documents, with a limit on the searching time of ten minutes. In these searches less than 0% of the assumed relevant documents were retrieved, but the precision ratio was 76%.

The search times have been recorded and are entered on the various record sheets. It is doubtful if the pre-search times can be compared, since different methods were used by the Centres, but there appears to be no correlation between pre-search and terminal times. The average terminal times at the Centres are given in Table 5, and for the four search modes in Table 6, and these latter show no significant differences.

	Search time (minutes)
Centre 1	37
Centre 2	42
Centre 4	29
Centre 5	57

Table 5. Average Centre Search Time

	Search Time (minutes)
Natural Language	37
Natural Language and A.C. File	45
Controlled Language	38
Controlled Language & Natural Language	40

Table 6 Average Mode Search Time

Question	Searcher	Total Ret.	Relevant			Estimated Relevant			Search Times		Precision		Matched Recall		Base Recall		Title Retrieval		
			1	2	X	1	2	X	Pre	Term.	1	1 & 2	1	1 & 2	1	1 & 2	1	2	X
NATURAL LANGUAGE																			
1 -02	B	18	4	7	7	4	7	7	30	47	22	61	100	91	100	100	4	1	1
-04	B	44	4	4	2	17	18	9	30	41	40	80	71	62	50	66	3	2	0
-06	A	90	0	2	18	0	9	81	10	40	0	10	0	75	0	50	0	0	2
-08	A	18	4	4	10	4	4	10	35	12	22	44	100	100	100	100	4	3	5
-10	A	18	4	5	0	8	10	0	0	22	44	100	66	50	0	25	2	1	0
NATURAL LANGUAGE & ASSOCIATED CONCEPT FILE																			
1 -01	B	49	0	1	9	0	5	44	15	28	0	10	0	25	–	–	0	0	1
-03	B	61	1	6	10	4	21	36	10	53	6	41	100	93	100	100	0	0	0
-05	B	2	1	1	0	1	1	0	10	70	50	100	50	66	100	50	0	0	0
-07	A	25	5	3	8	8	5	12	45	37	31	50	100	82	100	100	5	1	0
-09	A	30	2	4	10	4	8	18	30	18	12	38	100	100	100	100	0	2	3
CONTROLLED LANGUAGE																			
1 -02	A	17	4	4	9	4	4	9	30	37	24	47	100	66	100	100			
-04	A	43	3	5	2	13	21	9	30	45	30	80	100	87	50	66			
-06	B	1	1	0	0	1	0	0	10	49	100	100	100	25	100	50			
-08	B	4	2	1	1	2	1	1	35	20	50	75	50	37	33	28			
-10	B	27	6	8	0	11	16	0	0	23	43	100	66	68	66	75			
CONTROLLED LANGUAGE & NATURAL LANGUAGE																			
1 -01	A	76	0	3	12	0	15	61	30	37	0	20	0	100	–	–			
-03	A	23	0	6	2	0	17	6	20	17	0	75	50	69	100	100			
-05	A	158	1	0	19	8	0	150	30	50	5	5	100	66	100	100			
-07	B	15	4	2	4	6	3	6	45	54	40	60	100	82	100	100			
-09	B	16	1	6	3	2	9	5	30	24	10	70	50	80	100	100			

Fig. 10a ESTEC Summary Record Sheet

Question	Searcher	Total Ret.	Relevant			Estimated Relevant			Search Times		Precision		Matched Recall		Base Recall		Title Retrieval		
			1	2	X	1	2	X	Pre.	Term.	1	1 & 2	1	1 & 2	1	1 & 2	1	2	X
NATURAL LANGUAGE																			
2-02	B	5	0	3	2	0	3	2	5	39	0	60	-	100	–	100	0	0	0
-04	B	21	3	8	10	3	8	10	10	44	14	52	100	100	–	100	1	0	0
-06	A	41	4	3	4	15	11	15	0	45	36	64	45	44	0	0	0	0	0
-08	A	37	3	9	7	6	17	14	0	25	16	63	100	84	100	100	2	2	1
-10	A	30	8	9	2	12	14	4	0	15	42	90	100	100	50	50	4	3	0
NATURAL LANGUAGE & ASSOCIATED CONCEPT FILE																			
2-01	B	16	1	5	4	2	8	6	0	85	10	60	50	54	–	–	0	0	0
-03	B	98	7	5	6	38	27	33	0	65	39	67	100	100	100	83	4	1	0
-05	B	98	2	2	7	18	18	62	0	41	18	36	0	90	100	50	0	0	0
-07	A	27	1	2	7	3	6	18	0	21	10	30	25	25	50	60	0	1	0
-09	A	88	0	2	10	0	14	74	0	25	0	16	100	80	–	–	0	0	0
CONTROLLED LANGUAGE																			
2-02	A	2	0	2	0	0	2	0	5	7	0	100	–	67	–	0			
-04	A	8	2	1	5	2	1	5	10	25	25	37	67	27	–	0			
-06	B	53	9	4	1	34	15	4	0	30	64	93	90	82	0	33			
-08	B	10	1	5	1	2	6	2	0	57	14	85	25	35	33	33			
-10	B	6	5	1	0	5	1	0	0	49	83	100	45	30	25	25			
CONTROLLED LANGUAGE & NATURAL LANGUAGE																			
2-01	A	88	2	4	10	12	22	54	5	29	13	38	100	63	–	–			
-03	A	8	6	2	0	6	2	0	5	35	75	100	60	54	60	50			
-05	A	96	2	4	6	16	32	48	10	42	17	50	50	70	0	0			
-07	B	36	7	3	3	20	8	8	10	19	54	77	100	91	100	100			
-09	B	76	1	2	8	7	14	55	0	45	9	18	100	100	–	–			

Fig. 10b Technology Reports Centre Summary Record Sheet

Question	Searcher	Total Ret.	Relevant			Estimated Relevant			Search Times		Precision		Matched Recall		Base Recall		Title Retrieval		
			1	2	X	1	2	X	Pre.	Term.	1	1 & 2	1	1 & 2	1	1 & 2	1	2	X
NATURAL LANGUAGE																			
4-02	B	44	1	3	10	3	10	31	30	40	7	29	50	71	100	100	1	2	0
-04	B	6	0	3	3	0	3	3	25	30	0	50	0	60	–	–	0	0	0
-06	A	8	5	1	0	7	1	0	20	27	83	100	55	37	0	0	1	0	0
-08	A	32	6	1	12	11	2	19	12	42	33	38	100	88	100	100	3	0	0
-10	A	18	6	3	0	12	6	0	10	14	67	100	70	68	66	66	2	1	0
NATURAL LANGUAGE & ASSOCIATED CONCEPT FILE																			
4-01	B	53	1	9	1	5	43	5	20	40	9	91	100	90	–	100	1	5	0
-03	B	32	2	4	5	6	12	14	30	50	18	54	100	60	0	0	1	1	0
-05	B	20	0	2	18	0	2	18	15	25	0	10	–	100	–	100	0	2	1
-07	A	19	6	5	2	9	7	3	5	11	46	85	85	78	100	100	6	3	2
-09	A	108	9	4	2	65	29	14	12	30	60	87	91	94	50	50	5	2	1
CONTROLLED LANGUAGE																			
4-02	A	15	1	2	4	2	4	9	18	30	14	43	50	57	0	0			
-04	A	3	1	1	1	1	1	1	12	38	33	66	100	40	–	–			
-06	B	22	4	6	6	6	8	8	20	24	25	62	44	62	0	0			
-08	B	14	3	1	5	5	2	7	35	32	33	44	67	63	100	100			
-10	B	37	7	4	4	17	10	10	15	26	48	74	90	94	66	66			
CONTROLLED LANGUAGE & NATURAL LANGUAGE																			
4-01	A	42	3	6	1	13	25	4	18	30	30	90	75	68	–	100			
-03	A	30	0	4	6	0	12	18	12	48	0	40	0	40	0	0			
-05	A	2	0	1	1	0	1	1	10	12	0	50	–	50	–	100			
-07	B	21	3	3	8	5	5	11	25	20	21	43	43	43	100	100			
-09	B	55	3	3	2	21	21	13	25	22	38	75	54	66	25	25			

Fig. 10c Lund University Summary Record Sheet

Question	Searcher	Total Ret.	Relevant			Estimated Relevant			Search Times		Precision		Matched Recall		Base Recall		Title Retrieval		
			1	2	X	1	2	X	Pre.	Term.	1	1 & 2	1	1 & 2	1	1 & 2	1	2	X
NATURAL LANGUAGE																			
5 -02	B	37	6	9	3	12	19	6	5	74	33	83	87	95	66	75	3	5	0
-04	B	2	1	0	1	1	0	1	0	43	50	50	100	100	100	100	1	0	1
-06	A	4	3	1	0	3	1	0	10	67	75	100	75	44	—	50	2	1	0
-08	A	47	0	14	4	0	36	11	54	38	0	78	100	85	100	100	0	4	0
-10	A	76	4	0	9	24	0	52	15	17	30	30	71	75	100	100	3	0	0
NATURAL LANGUAGE & ASSOCIATED CONCEPT FILE																			
5 -01	B	8	1	2	5	1	2	5	15	115	12	38	33	30	-	-	0	0	0
-03	B	16	2	3	11	2	3	11	10	73	12	32	67	55	-	-	0	0	0
-05	B	44	7	5	10	14	10	20	15	48	32	54	100	100	-	100	3	1	0
-07	A	21	4	2	0	14	7	0	5	30	67	100	71	63	100	80	4	1	0
-09	A	46	2	3	7	8	12	26	10	25	16	42	100	81	100	100	1	2	0
CONTROLLED LANGUAGE																			
5-02	A	12	2	2	2	4	4	4	15	35	33	67	25	21	33	50			
-04	A	1	1	0	0	1	0	0	15	36	100	100	100	100	100	100			
-06	B	11	1	4	6	1	4	6	45	77	11	45	25	55	—	50			
-08	B	15	1	6	0	2	13	0	41	15	14	100	100	43	100	100			
-10	B	58	4	1	6	21	5	32	10	74	36	45	85	87	33	40			
CONTROLLED LANGUAGE & NATURAL LANGUAGE																			
5-01	A	14	2	6	6	2	6	6	30	10	14	57	67	80	—	—			
-03	A	7	1	3	3	1	3	3	0	93	14	57	33	47	—	—			
-05	A	16	5	2	2	9	4	3	5	64	55	78	71	54	—	100			
-07	B	81	5	2	10	24	9	48	45	93	30	42	100	100	100	100			
-09	B	54	1	6	6	4	25	25	0	79	8	54	100	100	100	100			

Fig. 10d Royal Institute of Technology Summary Record Sheet

Mode	Total Retrieved	Assessed Relevant 1	2	X	Assumed Relevant 1	2	X	Titles Relevant 1	2	X
N.L	188	16	22	37	33	48	107	13	7	8
N.L + A.C. File	167	9	15	37	17	40	110	5	3	4
C.L	92	16	18	12	31	42	19			
C.L. + N.L.	288	6	17	40	16	44	228			

Mode	Average Search Time Pre-.	Term	Precision 1	1+2	Matched Recall 1	1+2	Base Recall 1	1+2	Aver. of Numbers Precision 1	1+2
N.L.	16	32	28	59	67	76	50	68	18	43
N.L + A.C. File	22	46	20	48	70	73	100	88	10	34
C.L.	21	35	49	80	83	56	70	64	34	79
C.L. + N.L.	31	36	11	46	60	77	100	100	6	21

Fig. 11a Summary Results Sheet for ESTEC

Mode	Total Retrieved	Assessed Relevant 1	2	X	Assumed Relevant 1	2	X	Titles Relevant 1	2	X
N.L.	134	18	32	25	36	53	45	7	5	1
N.L. and A.C. File	327	11	16	34	61	73	193	4	2	0
C.L.	79	17	13	7	43	25	11	—	—	—
C.L. and N.L	304	18	15	27	61	78	165	—	—	—

Mode	Average Search Time Pre-.	Term	Precision 1	1+2	Matched Recall 1	1+2	Base Recall 1	1+2	Aver. of Numbers Precision 1	1+2
N.L.	3	34	22	66	86	86	50	70	24	67
N.L. and A.C. File	0	47	15	42	75	70	50	64	19	41
C.L.	3	34	37	83	57	48	19	18	46	80
C.L. and N.L.	6	34	34	56	82	76	53	50	20	46

Fig. 11b Summary Results Sheet for Technology Report Centre

Mode	Total Retrieved	Assessed Relevant 1	2	X	Assumed Relevant 1	2	X	Titles Relevant 1	2	X
N.L.	108	18	11	25	33	22	53	7	3	0
N L + A.C. File	232	18	24	28	85	93	54	13	13	4
C.L.	91	16	14	20	31	25	35			
C.L. + N.L.	150	9	17	18	39	64	47			

Mode	Average Search Time Pre-.	Term	Precision 1	1+2	Matched Recall 1	1+2	Base Recall 1	1+2	Aver. of Numbers Precision 1	1+2
N.L.	22	33	38	63	55	65	62	60	33	53
N.L. + A.C. File	16	31	27	65	94	84	50	70	37	77
C.L.	19	27	31	58	70	63	41	41	32	60
C.L. + N.L.	18	26	18	60	41	52	41	65	26	69

Fig. 11c Summary Results Sheet for Lund University

Mode	Total Retrieved	Assessed Relevant 1	2	X	Assumed Relevant 1	2	X	Titles Relevant 1	2	X
N.L.	166	14	24	17	40	56	70	9	10	1
N.L. and A.C. File	135	16	15	33	39	34	62	8	4	0
C.L.	97	9	13	14	29	26	42			
C.L. and N.L.	172	14	19	27	40	47	85			

Mode	Average Search Time Pre-.	Term	Precision 1	1+2	Matched Recall 1	1+2	Base Recall 1	1+2	Aver. of Numbers Precision 1	1+2
N.L.	17	48	38	68	87	80	91	85	24	58
N.L. and A.C. File	11	58	28	53	74	66	100	93	29	53
C.L.	25	57	39	72	67	61	67	68	30	57
C.L. and N.L.	16	64	24	58	74	76	100	100	23	52

Fig. 11d Summary Results Sheet for Royal Institute of Technology

Centre	Total Retrieved	Assessed Relevant 1	2	X	Assumed Relevant 1	2	X	Titles Relevant 1	2	X
1	188	16	22	37	33	48	102	13	7	8
2	134	18	32	25	36	53	45	7	5	1
4	108	18	11	25	33	22	53	7	3	0
5	166	14	24	17	40	56	70	9	10	1

	Average Search Time Pre-.	Term	Precision 1	1+2	Matched Recall 1	1+2	Base Recall 1	1+2	Aver. of Numbers Precision 1	1+2
1	16	32	28	59	67	76	50	68	18	43
2	3	34	22	66	86	86	50	70	24	67
4	22	33	38	63	55	65	62	60	33	53
5	17	48	38	68	87	80	91	85	24	58

Fig. 12a Summary Results Sheet for Natural Language (s.e. $\pm$ 6%)

Centre	Total Retrieved	Assessed Relevant 1	2	X	Assumed Relevant 1	2	X	Titles Relevant 1	2	X
1	167	9	15	37	17	40	110	5	3	4
2	327	11	16	34	61	73	193	4	2	0
4	232	18	24	28	85	93	54	13	13	4
5	135	16	15	33	39	34	62	8	4	0

Centre	Average Search Time Pre-.	Term	Precision 1	1+2	Matched Recall 1	1+2	Base Recall 1	1+2	Aver. of Numbers Precision 1	1+2
1	22	46	20	48	70	73	100	88	10	34
2	0	47	15	42	75	70	50	64	19	41
4	16	31	27	65	94	84	50	70	37	77
5	11	58	28	53	74	66	100	93	29	53

Fig. 12b Summary Results Sheet for Natural Language and Concepts (s.e. ±6%)

Centre	Total Retrieved	Assessed Relevant 1	2	X	Assumed Relevant 1	2	X	Titles Relevant 1	2	X
1	92	16	18	12	31	42	19			
2	79	17	13	7	43	25	11			
4	91	16	14	20	31	25	35			
5	97	9	13	14	29	26	42			

Centre	Average Search Time Pre-.	Term	Precision 1	1+2	Matched Recall 1	1+2	Base Recall 1	1+2	Aver. of Numbers Precision 1	1+2
1	21	35	49	80	83	56	70	64	34	79
2	3	34	37	83	57	48	19	18	46	80
4	19	27	31	58	70	63	41	41	32	60
5	25	57	39	72	67	61	67	68	30	57

Fig. 12c Summary Results Sheet for Controlled Language (s.e. ± 6%)

Centre	Total Retrieved	Assessed Relevant 1	2	X	Assumed Relevant 1	2	X	Titles Relevant 1	2	X
1	288	6	17	40	16	44	228			
2	304	18	15	27	61	78	165			
4	150	9	17	18	39	64	47			
5	172	14	19	27	40	47	85			

Centre	Average Search Time Pre-.	Term	Precision 1	1+2	Matched Recall 1	1+2	Base Recall 1	1+2	Aver. of Numbers Precision 1	1+2
1	31	36	11	46	60	79	100	100	6	21
2	6	34	34	56	82	76	53	50	20	46
4	18	26	18	60	41	52	41	65	26	69
5	16	64	24	58	74	76	100	100	23	52

Fig. 12d Summary Results Sheet for Controlled Language and Natural Language (s.e. $\pm$ 6%)

COST–EFFECTIVENESS

The two measures of recall and precision have been criticised on the grounds that they are twin measures and that they do not present a single absolute measure. One such measure that has been used is the 'normalized recall ratio' but this requires a ranked out-put of documents, and is certainly not applicable in this test. Although various other single measures have been proposed, these have tended to be the proposals of mathematicians and none proved workable or meaningful in practice.

The use of recall ratio in a test of an operational system can be criticised on various grounds but in this test it is the precision ratio which causes more concern. As originally devised and previously used, it has been based on the total number of documents retrieved. In a non-interactive search mode this is a single definite figure. In this test, with complex interactive searching, many documents may be 'retrieved' in the course of the search, but the final print-out may, by the decision of the searcher, be only a sub-set of these 'retrieved' documents. It is the documents in the final print-out which are considered for the calculation of the precision ratio, and a check on the figures in the Centre Record Sheets will show that this appears in some cases to be an arbitrary figure; for the same question one searcher has 100 items in the final print-out, the other searcher, having consulted in the course of the search the titles or abstract of many documents, finally retrieves only a single item.

One measure that overcomes most of the problems is a measure that not only includes both recall and precision, but also takes account of time and introduces the element of cost. This cost-effectiveness measure calculates the cost of retrieving a single relevant document, and is normally expressed as :

$$C_r = \frac{C + F\left\{\left(D \times \frac{R}{100}\right)\left(\frac{100}{P}\right) - \left(D \times \frac{R}{100}\right)\right\}}{D \times \frac{R}{100}} = \frac{100C}{DR} + \frac{100F}{P} - F$$

where C = system costs of a single search
D = expected number of relevant documents
R = recall ratio
P = precision ratio
F = charge for non-relevant citations

F, the charge or 'fine' for non-relevant documents, is required to differentiate between a search where, for example, 10 relevant documents are retrieved on their own as against a search where 10 relevant and 100 non-relevant documents are retrieved. For the purpose of this test, this measure can be simplified to:

$$C_r = \frac{C_s + (F \times D_n)}{D_r}$$

where C_s = costs of a single search
D_r = number of relevant documents retrieved
D_n = number of non-relevant documents retrieved
F = fine for retrieving a non-relevant document

The cost of an on-line search will be dependent on:

1. Service charges
2. Line charges
3. Printout charges
4. Operator's time cost
5. Customer's time cost in assisting in search or in rejecting nonrelevant documents.

These are variables in any given situation, but generally speaking in a normal on-line search items 1, 2, 4 and 5 are time dependent. Item 3 is an added charge if off-line printing is used, if the printout is on-line, then it is also time dependent.

One can hypothesise a number of different situations and calculate the effects of varying the costs of any of the five aspects listed above. For the use of this measure in a given operational system, actual costs should be obtained, but for comparative purposes, as in this test, it is not essential to have an exact figure, and a reasonable figure of $2 a minute connect-time will be taken as the combined search costs and the 'fine' will be taken to be $0.10.

To illustrate with an example, for search 1.07 the estimated number of documents retrieved and the search times for the two modes were:

	Relevance 1 and 2	Non-relevant	Search time (minutes)
N.L.	13	12	37
C.L.	9	6	54

Therefore for N.L.

$$C_r = \$ \frac{(2 \times 37) + (12 \times 0.1)}{13} = \$5.78$$

and for C.L.

$$C_r = \$ \frac{(2 \times 54) + (6 \times 0.1)}{9} = \$12.07$$

C_r for the four Centres is shown in Table 7 with the final line giving the overall figure for the four search modes.

	N.L.	N.L. + A.C. File	C.L.	C.L. + N.L.
Centre 1	$4.08	$8.26	$4.82	$6.38
Centre 2	3.86	3.64	5.01	2.56
Centre 4	5.25	1.76	5.78	2.57
Centre 5	5.07	8.02	10.43	7.45
Average	$4.56	$5.42	$6.51	$4.74

Table 7 Cost-effectiveness figures

One should not read too much into the actual figures; apart from other factors, most of the searchers expressed the opinion that the requirement for keeping records of the search had increased their search times by up to 30%. Nor was there ever any suggestion in the test design that the searchers should do anything except attempt to optimise the performance within the limits of the particular search mode being used, and in this sense 'optimise' did not necessarily mean 'maximise' the number of relevant documents retrieved. This cost-effectiveness measure strongly emphasises recall; one has to raise the 'fine' to an absurd level before it seriously affects the final figure. In other words, from a cost-effectiveness viewpoint, the precision ratio would appear to be less important than recall ratio. However, the comparison between the four search modes is of interest, and even if it is of limited relevance in this test, the measure does appear to be of particular value in the evaluation of on-line systems where costs can be relatively easily assessed.

FAILURE ANALYSIS

An analysis was made of the two search outputs for each question to attempt to ascertain why one system had failed to retrieve a relevant document which the other system had retrieved. A number of different causes can be identified.

1. An apparent mismatch between the question and the documents. A paper might be retrieved, for example, by Natural Language through some accidental relationship of search terms, while C.L. search strategy was perfectly adequate but would not have retrieved the documents. The reasons for the questioner deciding that the document was relevant might not appear obvious to a third person. Such mismatches tended to occur with Relevance 2 rather than Relevance 1 documents.

2. An input failure which could be of two types:

a. an index failure, where a concept which is included in the title or abstract has not been included as an index term.

b. an abstract failure, where the full document presumably includes a given concept since it has been given as an indexing term, but which is not mentioned in the abstract. Alternatively a term might be misspelled or appear in an unusual form.

3. A search failure, of which there are, in theory, at least two types.

a. a searcher may fail to include all the appropriate terms in the search

b. the searcher may have required inappropriate or over-specific combinations of terms.

In practice it was often difficult to distinguish these two types, partly because of the extreme complexity of many of the searches, but also because it was a purely subjective decision as to whether it would have been better to include additional search terms as alternatives, or to weaken the search requirements by making it less specific.

4. Data base failures. Some cases were noted where there was no apparent reason why a relevant document should not have been retrieved, and one can only reach the conclusion that it was a data base failure.

Figure 13 illustrates in detail the method by which this analysis was done; the complete set of comments and assessments is included in Appendix E

The reasons for failure are summed in Table 8

	N.L.		N.L. + A.C.		C.L.		C.L. + N.L.	
	Rel. 1	Rel. 2	Rel. 1	Rel. 2	Rel. 1	Rel. 2	Rel.1	Rel.2
Searching	19	25	13	25	14	38	18	28
Indexing					18	19	3	6
Abstracts	2	6	0	3				
Mismatch	0	5	2	7	2	12	2	4
Data Base	2	3	0	0	1	0	2	0

Table 8. Summary of causes of recall failures

Q1–01

GENERAL COMMENTS: A difficult question. User did not know of any relevant papers, the manual search found nothing that was relevant and no Relevance 1 documents were found in either search. In spite of the two searches retrieving 76 and 49 references, only 6 were in common.

FAILURE ANALYSIS: All recall failures were due to mismatch of question and document as for example:

434944 deals with fatigue resistance and not corrosion, with 'welded' not 'brazed'.

333440 deals with fracture of pressure vehicles.

314565 has no mention of joints, and deals with stress corrosion.

ASSESSMENT : N.L. and
Rel. 2 Mismatch (3)

Q1–02

GENERAL COMMENTS: Straightforward question with appropriate C.L. terms. No Relevance 1 failures.

FAILURE ANALYSIS: All failures were of Relevance 2 documents, as following examples:

435128 Indexing does not mention fibre optics

422325 It is stated in the abstract that the system is suitable for transmission of information, but this concept is not included in the index terms.

421037 Paper deals with interferometer ring dissector

323237 Fibre optics mentioned in abstract but not indexed.

324199 No mention of fibreglass in abstract.

ASSESSMENT: C.L. N.L.
Rel. 2 Indexing (4)
Rel. 2 Abstract (1)

Fig. 13 Failure Assessment Sheet

These figures are of interest, but must be treated with caution, and certainly the apparent preponderance of failures assigned to searching should not be taken as an indication of incompetence on the part of the searchers. For example, many search failures in Natural Language were due to the searcher not having included all variant word endings (e.g. trail, trails, trailing), but it could be argued that this was due to the system as, being experimental, it had not been equipped with the normal facility for confounding word endings. On the other hand, many search failures in Controlled Language were due to a failure to include all possible index terms; here a more detailed analysis might indicate that there were over-subtle distinctions between many of the controlled language terms. In making these assessments of recall failures, it is also necessary to appreciate that changing a search strategy so as to retrieve missing relevant documents would also have resulted in the retrieval of many more non-relevant documents. Reference to the record sheets maintained by the searchers leads to the subjective conclusion that possibly the majority of the searching failures were the result of a deliberate decision to restrict the retrieval of non-relevant papers.

From the Master Record Sheets can be obtained an analysis of the overlap, that is, the number of documents retrieved by both search for the same question. To summarise this as a simple number would be inappropriate, since the overlap depends both on the lower number of documents retrieved by either search and on the number retrieved by the other search. For example, if one search retrieves only four references, there cannot possibly be more than four in the overlap. However, the probability of this would be greater if the other search had retrieved 100 documents than if it had also only retrieved four documents.

To take these two factors into consideration, the percentage overlap is expressed as:

$$P_O = 100 \frac{L}{M} \times \frac{L}{D}$$

where L = overlap
M = minimum number of documents retrieved by either search
D = total of different documents retrieved by both searches.

This equation would, for a perfect match, give 100%, but as a more normal example, assume a search where there is an overlap of ten items, with one search retrieving fifteen documents and the other search retrieving forty documents. Therefore, D would represent:

$$(15 + 40) - 10 = 45$$

$$P_O = 100\left(\frac{10}{15} \times \frac{10}{45}\right) = 14.5$$

P_O for the searches at each Centre are given in Table 9. There does not appear to be any particular reason for the major difference in the average figure for Centre 4. It might be expected that there would be a significant difference between those searches where a comparison was between Controlled Language and Natural Language (even-numbered questions) and those where comparison was between C.L + N.L. and N.L + Concept (odd-numbered questions). Because both the latter modes had been mainly N.L. terms, the hypothesis was that they would have the higher score. In fact the reverse is the case, the averages being 12.1 for the former and 11.0 for the latter.

	Centre	1	2	4	5
Question	01	0.5	1.0	12.7	3,5
	02	23.6	40.0	1.1	5.6
	03	22.0	8.1	0	1.0
	04	2.3	13.1	0	50.0
	05	1.2	23.2	2.5	31.0
	06	0	8.2	0	0
	07	17.4	4.3	3.7	22.9
	08	22.2	3.7	9.0	7.9
	09	31.6	16.6	13.2	5.7
	10	2.2	20.0	18.4	16.1
Average		12.3	13.8	6.0	14.7

Table 9 Overlap ratio for each question by Centres

In a recent paper Bourne (Ref.6) analysed the spelling errors that are found in various data bases and considered their effect on performance. He found considerable variation amongst data-bases, ranging from 22.8% in Abstracted Business Information down to 0.4% in BIOSIS, with the frequency of posting to a misspelled term ranging from one posting in 160 citations to one in 8,000. His conclusion is that these misspellings 'have relatively little impact on file use for many data bases'.

The NASA data base was not included in his analysis, but in in the present test, only one case of a spelling variation was found to have been the cause of a search failure, and that occurred in the abstract (cesium for caesium).

As a by-product of his work, Bourne has also analysed the overlap of index terms occurring in eleven data bases. He found that the sum of 5,898 terms in separate lists would be reduced to 3,608 shared terms. Of these shared terms, 2,736 occurred in only one of the eleven separate lists.

With the other modes retrieving more documents than C.L., it was to be expected that more relevant documents - and thereby a higher recall ratio - would be retrieved. However, as considered earlier (page 22) the method adopted for determining the matched recall ratio was one which had a tendency to improve the figures for the system which retrieved fewer documents. Had Method 3 been adopted, the comparative performance would have been significantly better for those search modes which consisted of or included N.L. searching.

A further analysis that can be made compares the number of relevant documents retrieved by the other modes for each relevant document retrieved by C.L., and this is shown in Table 11.

DISCUSSION OF RESULTS

To make an impartial analysis of an evaluation of an information retrieval system is a difficult task. The results are rarely conclusive, and somewhere in the results there can usually be found supporting evidence for diametrically opposed viewpoints. Given that system A has a recall and precision ratio both of 70%, there would usually be agreement that it was performing better than system B with a recall and precision both of 50%. However, such unambigious examples are rare, and it depends on a personal viewpoint whether a recall of 70% with precision of 50% is better or worse than when both recall and precision are 60%.

While the difficulty still exists in this case, it is made somewhat easier because of the limited objectives of the test. The major objective of the test was to obtain data to provide an answer to the question whether in an operational system natural language searching on titles and abstracts could match the performance obtained with conventional searching on controlled language index terms. In whatever way the results are considered it appears, from the available evidence from this test that natural language searching is, to say the least, capable of comparable performance.

If there is one set of data that qualifies this statement, it relates to the number of documents retrieved in the various search modes. In Table 10 is shown the number of documents retrieved in the other search modes for each document retrieved by Controlled Language at the Centres.

	Controlled Language	Natural Language	N.L. and Concepts	C.L. and N.L.
Centre 1	1	2.04	1.81	3.13
2	1	1.69	4.13	3.84
4	1	1.76	3.22	2.08
5	1	1.71	1.39	1.77
Average		1.80	2.63	2.70

Table 10 Comparison of total document retrieval by four search modes

With the other modes retrieving more documents than C'L. it is to be expected that more relevant documents - and thereby a higher recall ratio - will be retrieved. A further analysis that can be made compares the number of relevant documents retrieved by the other modes for each relevant document retrieved by C L., and this is shown in Table 11

		Controlled Language	Natural Language	N.L. and A.C. File	C.L. and N.L.
Centre	1	1	1.11	0.78	0.72
	2	1	1.31	1.97	2.04
	4	1	1.47	3.78	2.19
	5	1	1.74	1.33	1.58
Average			1.41	1.96	1.63

Table 11. Comparison of known relevant document retrieval by four search modes.

The figures underlined in Table 11 indicate the three cases in which the the proportion of relevant documents retrieved is greater that the proportion of the total documents retrieved as given in Table 10. Although overall no search mode achieves a match between the two figures, Natural Language has the most favourable comparison with 1.41 as against 1.80. The searches by the two modes of Controlled Language and Natural Language were made on the same questions, so a direct comparison can be made between them. These figures show that 51% of the additionally retrieved documents were relevant, and considering these came at what might be termed the tail end of the retrieval, this represents a reasonable precision ratio.

If the averages of relevant documents retrieved (Table 11) are expressed as percentages of the average of total documents retrieved (Table 10), these are 78% for N.L., 74% for N.L. & A.C.File and 60% for C.L. and N.L. The lower figure for the latter is partly explained by the fact that the overall recall ratio was particularly affected by the results in this mode at Centre 4. Given the option of using controlled language or natural language, the other centres made extensive use of natural language terms, but at Centre 4 the searches were almost entirely controlled language index terms. For this reason it is valid to present the results for Centre 4 grouping the two Natural Language modes and the two Controlled Language modes. In such a case the performance is as in Table 12.

	Recall	Precision
Natural Language	74%	64%
Controlled Language	58%	59%

Table 12 Comparative performance for searches at Lund University (s.e. ± 6%)

While these results are based on only 10 searches, they do indicate a probably superior performance for Natural Language.

This is not a surprising result. Although it was not appreciated fully at the time, the phenomenon of the effectiveness of single term natural language searching was first demonstrated in the ASTIA-UNITERM test in 1953 (Ref. 7) and later in the results of Cranfield 1 (Ref. 8) and particularly in the tests carried out by Swanson. (Ref. 9) It was finally proved, at least in regard to small test collections, in Cranfield 2 (Ref. 4) and has subsequently been verified in a number of other test situations (e.g. Ref. 10).

There appear to be two matters which significantly affect the performance of information retrieval systems, with a third aspect that might in theory do so, but which rarely appears to in practice. This latter point refers to the quality of the indexing or the surrogate on which a search is based. Rank bad indexing could be described as assigning an index term which is completely unrelated to the subject of the document, as, for example, if a document dealing with satellite orbits were indexed with the terms 'milling machines' or 'sheep grazing'. Needless to say, such absurd indexing never happens in reputable organisations. However, in any test of any index, it is very easy to find examples of indexing which can be described as less than perfect. In the two main Cranfield tests, indexing errors accounted for 30% and 59% of the search failures; in the Mediars evaluation the figure was 37% (ref. 11) In this test (see Tab.8) it was 28% which appears to be about a level which one has to accept.

Even in the best systems, occasional human errors can be made (e.g. macro-climate instead of micro-climate) and will escape the most rigid controls, but most of the 'errors' of indexing are created by a management decision to impose a limit on the level of exhaustivity of indexing, which is the most important factor affecting the recall performance of I.R. systems. To consider the example mentioned earlier (Ref. 2) two papers, identical word-for-word, were indexed at different times. In the first instance, twelve terms were assigned, of which five were major terms. In the second instance fifteen terms were assigned, and again five were major terms, Four of the major terms were common to the two papers, but only two minor terms were common. The first paper had six terms not used to index the second paper, which in turn had nine terms not used to index the first. As to which set of index terms were the better would be entirely dependent on the question being put to the system. One could easily hypothesise one question retrieving the paper on the first set of index terms but not on the second, and another question where the position would be reversed. The only way both questions could have retrieved the paper would be to index by the complete set of 21 terms. This is a simple illustration of the major effect which exhaustivity has on recall, and it is undeniable that failure to index a concept is certain to result in failure to retrieve the item by that concept.

The other aspect which is of importance is the specificity of the index language. If the index terms are so broad that many differing concepts are lumped together, the precision will be adversely affected. If the index terms are over-precise, then it is probable that recall will suffer.

Experimental work has shown that there is an optimum level for exhaustivity of indexing and specificity of the index language in any given situation, both in respect of each separately and also in conjunction. It appears that the higher the level of exhaustivity of indexing, the more specific can the index language be; a low level of exhaustivity requires a less specific index language.

However, the distinctions are quite fine, and can probably only be assessed in a controlled experiment. What is reasonably certain is that, within the levels of what has been here investigated (i.e. abstracts or NASA index terms) the higher exhaustivity is likely to give significant improvement in recall without too seriously affecting precision.

Therefore it is argued that it is the on-average greater exhaustivity of the abstract which is responsible for the improvement in recall ratio when compared to searches on index terms, and the theoretical improvement which should come from intellectual analysis of the document by skilled indexers is not sufficient to counterbalance the effect of exhaustivity.

A secondary objective of the test was to consider the effect on natural language of the use of the Associated Concepts File. This is primarily a device which it is intended should improve the recall ratio by giving to the searcher clues as to possibly useful additional search terms. As can be seen from Table 13, the only Centre wherethe use of the Associated Concept File resulted in a higher recall performance than that of Natural Language on its own was at Centre 4, whereas at the other three Centres recall was somewhat inferior , as it was with the average of all searches.

	Natural Language & A.C. File		Natural Language	
	Recall	Precision	Recall	Precision
Centre 1	73	48	76	59
Centre 2	70	42	86	66
Centre 4	84	65	65	63
Centre 5	66	53	80	68
**Average	73	52	77	64

Table 13 Comparison of searches with and without Associated Concepts File (** s.e. $\pm$ 4%)

This cannot be taken as conclusive, since the searches were on different sets of questions, but also of some interest is the reaction of the searchers to the Associated Concepts File. At the conclusion of the test, the searchers were asked to express their agreement or disagreement (on a five-point scale) to a number of statements relating to the test. One such statement was

> 'With natural language searching an auxiliary aid, such as the associated concept file, is essential.'

Of the six replies received, two searchers at Lund were in agreement,

but the remaining four strongly disagreed. This bears out a subjective impression, derived from a study of the comments of the searchers on their strategy that in most cases the starting set of search terms was quite obvious, and that additional terms came from titles or abstracts of papers found in the search.

The complete questionnaire sent to the searchers is shown as Figure 14, and an analysis of the replies is given in Table 14

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Q1		1		3	2
Q2		2			4
Q3	1	3	2		
Q4	3	1	1	1	
Q5	3		2	1	
Q6				2	4

Table 14 Analysis of replies to questionnaire (Fig. 14)

While this is such a small sample as to have little claim to validity, it is of interest in reinforcing the overall test results. Changing, where appropriate, from negative to positive statements, it seems that, as a group, the searchers believe

1 that it is easier to cover all possibilities with natural language
2 that it is easier to match the user's requirements with natural language
3 that control of word endings is essential with natural language
4 that ability to use natural language and controlled language improved the search.

The searchers were also asked to express their views on the suitability of the NASA Data Base for each of their test questions, and to state, where appropriate, which other data base they would have used under normal operational conditions. The two searchers were in complete agreement at ESTEC in deciding that the NASA Data Base would have been their first choice for eight of the questions, with one search being on Metadex and one on Inspec. There was also agreement at Technology Reports Centre, with NASA Data Base being first choice for six searches, with two on Compendex and one each on CAC and INSPEC. However at Lund there were differing views and the first choice would have been

	Searcher A	Searcher B
NASA	4	2
INSPEC	3	4
COMPENDEX	2	2
N.S.A.	1	
CACON		2

COMMENTS ON NASA TEST

Searcher A

Please indicate against each of the statements below whether you agree, disagree or are neutral by putting a figure against each statement to mean the following:

1 - Strong agree

2 - Agree

3 - Neither agree or disagree

4 - Disagree

5 - Strongly disagree

1. I was more confident of having covered all possibilities with the controlled language search than with the natural language search. [5]

2. With natural language searching an auxiliary aid, such as the associated concept file, is essential. [5]

3. It is easier to match the questioners requirements with a free language search than with controlled language. [2]

4. I took longer to carry out each test search than would normally be the case.
(If there is agreement with this statement, please indicate by how much on average (e.g. 10% or 20% etc). [2] [10 %]

5. In an operational system using natural language it would be essential to control word endings (e.g. singular and plural etc). [4]

6. The freedom to use both controlled language and natural language did not appear to improve a search. [5]

Fig. 14 Completed copy of questionnaire for searchers

An analysis was made of the seven N.L. and Concept searches made at Estec and TRC where the original analysis showed that some relevant documents had not beenfound due to search failure (I-03, 1-04, 1-09, 2-03, 2-05, 2-07 and 2-09) The intention was to find whether any of these documents might have been retrieved by the use of Associated Concept File, since, from the comments made by the searchers at the two Centres, it appears that they were unlikely to have made much use of the A. C. File. Such analysis was difficult and highly subjective, but there appeared to be no search failures that could reasonably have been avoided by the use of the Associated Concept File.

An additional point in this respect is that one of the searchers at Lund, who had given a positive reply to the statement concerning the requirement for something such as the Associated Concept File, added the rider 'the Associated Concept File was not enough as an auxiliary aid, and a thesaurus, for instance the EJC thesaurus was essential.

CONCLUSIONS

It appears impossible to reach any other conclusion than that, within the parameters of this test, Natural Language searching on titles and abstracts proved at least equal to and probably superior to searching on Controlled Language terms. It also seems that a significant factor in this was the increased level of exhaustivity, and, therefore, there is no possibility that an equal level of performance could be attained by restricting the searches to titles.

While these conclusions are reached on a smaller test set than had been originally intended, there would appear to be no justification for suggesting that, given the same general type of search pattern, there would have been any significant difference in the result with a larger test set.

Availability of an Associated Concepts File (or a similar type of device) is something that is theoretically desirable, but which in this test did not appear to be particularly useful. It might be argued that its potential value was reduced by having been compiled from title terms only and by having been restricted to terms which occurred at least ten times. Certainly, in the form in which it was made available, it was very cumbersome to use, and it may be that, if it had been available on-line, the improved ease of reference would have resulted in its incresed use.

This leads into one caveat which I feel has to be made about the whole test, and this relates to the operating conditions. It was carried out from various Centres which had direct lines to Frascati and which also had VDU.s operating at 240 c.p.s. As a user who, for the past three years has been accessing on-line systems with a 10 c.p.s. teletype terminal and within the severe financial constraints of a library budget, the completely different approach adopted by the Centres was particularly striking. Obviously the Centres are correct in fully exploiting the interactive capabilities of on-line systems, whereas we have not been making full use of these because of the constraints imposed by the equipment and the cost. However, whatever might be the situation in the future, there is little doubt but that the majority of users as of now will be operating with, at best, 30 c.p.s. terminals and probably within cost restraints which would make many of the search times found in this test quite unrealistic. In such circumstances greater emphasis tends to be placed on the preparatory decisions before commencing the search, and it might be argued that this in turn implies that searching on Controlled Language indexing with the ability to consult a well-structured thesaurus is likely to be more effective than was the case in this test.

As an overall recommendation I think it best to quote the written comments of one of the senior searchers in this test:

" According to my experience a data base, always should offer both controlled and natural language. Controlled language will give an effective and quick result to questions suitable for the data base. On the other hand if problems arise to find exact descriptors you will be grateful to be able to use natural language. Very often the possibilities complement each other.'

REFERENCES

1. WALL, E. — Performance improvement effected by use of thesaurus. A.S.I.S. Proceedings. 1971, pp291-3

2. CLEVERDON, C. and J. KIDD — Redundancy, relevance and value to the user in the output of information retrieval syatems. Jnl. of Documentation, Vol. 32, 1976, pp159-173

3. CLEVERDON, C. — User evaluation of information retrieval systems Jnl. of Documentation, Vol. 30, 1974, pp170-180

4. CLEVERDON, C at al. — Factors determining the performance of indexing systems. Vol. 2. Cranfield. 1966

5. CLEVERDON, C. — On the inverse relationship of recall and precision. Jnl. of Documentation. Vol. 29, 1972, pp195-201

6. BOURNE, C. — Frequency and impact of spelling errors in bibliographic data bases. Information Processing and Management, Vol. 13, 1977 pp1-12

7. GULL, D.C. — Seven years work in the organisation of materials in the special library. American Documentation, Vol. 7,1956, pp320-329

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

9. SWANSON, D.R. — Word correlation and automatic indexing. Phase 1, Final Report. Ramo-Wooldridge. 1960

10. AITCHISON, T. et al. — Comparative evaluation of index languages. INSPEC Report R70/2, July 1970

APPENDIX A ORIGINAL TEST DESIGN

A COMPARATIVE EVALUATION OF SEARCHING BY CONTROLLED LANGUAGE AND NATURAL LANGUAGE IN THE NASA DATA BASE

Cyril W. Cleverdon

Cranfield Institute of Technology

REQUIREMENT

It is desired to investigate the comparative performance of on-line searches using either a controlled index language or a natural language based on terms in titles and abstracts. For this purpose a test file has been prepared from the full input to NASA Star for the years 1973 and 1974, consisting of approximately 50,000 references.

I TRODUCTION

In designing an evaluation of this type, there are certain limitations which have to be accepted. First, this is not an evaluation of an operational system, but is a test of a sub-system, namely the indexing and index language. Therefore it will not be concerned with many aspects which would have to be considered in a total evaluation of an operational retrieval system, such as the interface with the users, the consistency and accuracy of indexers or the expertise of those carrying out the search. Secondly, because it is a comparative test, it is essential to keep control of all the variables that are extraneous to the sub-system being tested. If this is not done it will, at the best, be necessary to increase significantly the size of the test and the scale of the analysis; at worst, an uncontrolled variable will be found to out-weigh the effect of the variable under test, so there will be a failure to meet the objectives of the test.

However, if the test is designed so that extraneous variables are controlled, there are some compensating advantages. A less complex methodology can be used, and somewhat cruder measures can be used. For example, a full evaluation of an operational system would not only require genuine questions but also that relevance judgements should be based on full text. However, in the experimental situation of this test, one can accept prepared questions and relevance judgements based on abstracts, doing this not because it is desirable but on the practical grounds that it simplifies the test methodology and will not have a significantly adverse effect on the comparative results.

There are, of course, some variables which cannot be controlled, and the most important of these is the search query. If all questions put to a system were of exactly the same kind, a satisfactory test would be possible using only one search. Since this is most obviously not the case, it is necessary to use sufficient questions as will reduce the effect of this variable, doing this in such a way that they are representative of the type of questions for which the system is designed.

TEST DESIGN

Seven user centres have agreed to take part in the initial stages of collecting test data, and each centre has accepted a commitment to process a minimum of ten questions. The test procedure will then be as follows.

1. The Project Director will prepare and supply to each centre detailed instructions and the necessary recording forms.
It is essential that the various centres should adhere strictly to the requirements of the test design, so that no unknown variables enter into the analysis of the test results. The Project Director will therefore prepare detailed instructions concerning the procedures to be followed in making the searches.

2. Each centre will recruit ten scientists or engineers who would agree to cooperate in the test.
While it would be most satisfactory if the test questions could be valid questions, this is not practic 1 within the proposed time scale. Therefore those cooperating will either be persons who have recently had a search made or who are willing to prepare a suitable question. Additionally they will be asked to supply a list of any papers (up to a maximum of six) of which they are already aware and which would be relevant to the search question. At a later date, they will also be required to assess the relevance of retrieved items. (see 5 below)

3. Separate searches will be made on the test collection using the controlled language and the natural language.
An original expectation with on-line systems was that the actual user would carry out his own searches, but in practice this does not appear to be the normal procedure, and searches are usually delegated to an intermediary. It will be the professional intermediaries at each centre who will be responsible for carrying out the test searches.
Since each search has to be duplicated, difficulties arise as to the procedure to be followed, bearing in mind the necessity of eliminating, as far as possible, uncontrolled variables. A search could, for example, be carried out quite independently by two different persons, but this could introduce not only the variable of the ability in searching of the two individuals but also the more serious variable of different interpretations of what the questioner really required. On the other hand, if the same person did both searches, the second search could be significantly influenced by the experience gained in the first search. Therefore there will be two intermediaries at each centre, each doing five searches with controlled language and five with natural language. However, before doing a search, the intermediaries will consider the question, where possible with the questioner himself, and will attempt to achieve an agreed interpretation of the meaning of the question, and also agree on a comparative and matching set of starting search terms.

Always a problem in this type of comparative test is the decision regarding the level of output. A search statement can be so structured that it will give low recall and high precision or one can go for high recall and low precision, these different levels of performance

being obtained by variations in the exhaustivity of the search, the specificity of the search terms or a combination of both. Unless there is some form of control, there will be the possibility in this test that one search may be stopped when ten documents have been retrieved but that the complementary search for the same question will be continued until one hundred documents are retrieved. From the viewpoint of the analysis, this would be an unsatisfactory position, to avoid which it will be necessary for the two searchers to reach prior agreement as to the level of the search and the extent to which the search can be broadened. Such agreement would, of course, only relate to the two searches for a single question, and the agreed level will probably vary between different searches at the same centre, and undoubtedly will vary between different centres.

In that a statistical analysis has been prepared of the words used in the titles of the documents in the test collection, it is required that the test show whether the use by searchers of this listing of associated concepts gives improved performance in the natural language searches. To obtain data on this point will require that half the natural language searches will be made with the operators being able to use the list of associated concepts to help in optimising the search strategy.

Apart from the requirements considered above, it is accepted that the search pattern will reflect the normal operating techniques of the searchers in the various centres. From discussion, it is clear that there are wide variations in the normal search pattern at different centres, and it would be impractible - and undesirable - to attempt to impose a rigid search procedure for all centres to follow. However, for the purpose of the analysis, it will be necessary for full records of the search pattern to be maintained, and this will be done, with the records annotated by the searchers when it is necessary to explain their reasons, for example, of a switch in search strategy.

4. A manual search will be made for each question in the appropriate section of NASA STAR for documents which appear to be relevant.
One method of obtaining the recall ratio is to base this on known relevant documents. It is expected that many of the questioners will be able to supply such known relevant documents, but it will probably be necessary to supplement what they are able to give. A search in a manual system is the simplest way of doing this, in that it is only necessary to find a few relevant items for each question. It might appear that it would introduce bias if these manual searches were made with NASA STAR, but the experience in the MEDLARS evaluation of using Index Medicus justifies an expectation that this would not be the case.

5. A maximum set of thirty citations will be obtained for relevance assessment.
The two searches on a given question having been completed, the print-outs will be sent to the Project Director. The outputs will be compared and the total set of citations will be ascertained. If this, plus any items located in the manual search, were less than thirty, the whole set would be sent for relevance decision by the questioner. If more than thirty, a sub-set of thirty, representative of the two outputs, would be prepared and sent for relevance assessment.

6 The questioner will decide the relevance of the citations
The questioner will receive a set of abstracts and, in relation to his question, will indicate for each item whether it is of major relevance, of minor relevance or non-relevant, and for all items in the final category indicate briefly the reason for his decision. The relevance decisions will be recorded on a standard form made available by the Project Director.

7 The search results will be analysed
The relevance decisions having been received by the Project Director, the results will be correlated for all the centres, and calculated in appropriate measures (for examples see Appendix A). Analysis will be made of search failures to ascertain the reason for any significant differences in the comparative performance, and to determine whether any recommendations can be made for overall improvement in performance.

8 Report
The Project Director will prepare and submit the final report by March 31st 1976

TIMING

The initial stages of obtaining agreement from scientists and engineers to cooperate will be started immediately. It is expected that the test data base will be ready for use by early December, during which month the searches will be made. It is hoped that the majority of the relevance assessments will be completed during January, allowing some two months for analysis and preparation of the report.

APPENDIX A

QUESTION ANALYSIS SHEET

Assume that, for a given question, C.L. retrieved 40 documents and N.L. retrieved 50 documents, and that 10 of them were retrieved only by C.L. and 20 only by N.L.
Therefore documents 1-10 are C.L., 11-40 are jointly C.L. and N.L. and 41-60 are N.L.

Assume relevance decisions on sample of 30 documents:

C.L.		C.L. and N.L.		N.L.	
1	R	11	R	41	R
3	N-R	13	N-R	43	R
5	N-R	15	R	45	N-R
7	N-R	17	R	47	N-R
9	R	19	N-R	49	R
		21	N-R	51	N-R
		23	R	53	R
		25	R	55	N-R
		27	N-R	57	N-R
		29	R	59	R
		31	R		
		33	N-R		
		35	R		
		37	N-R		
		39	N-R		
Totals	2R, 3N-R		8R, 7N-R		5R, 5N-R

Assume documents 11, 17 and 49 had been found beforehand by a manual search, plus one other document which was judged relevant but was not found in the mechanized search.

PERFORMANCE MEASURES

The usual measures are recall and precision, calculated on the following basis:

	Relevant	Non-relevant
Retrieved	a	b
Not retrieved	c	d

$$\text{Recall} = \frac{a}{a + c} \qquad \text{Precision} = \frac{a}{a + b}$$

Since the true figure for c cannot easily be determined, the alternative ways of calculating recall are:

Base recall $= \frac{a}{a + c}$ where a + c represents the total of relevant documents found outside the system being tested (i.e. manual search)

Matched recall $= \frac{a}{a + c}$ where a + c represents the total of relevant documents retrieved by the two systems being tested (i.e. C.L. and N.L.)

Therefore, the performance ratios for this question would be as follows:

	C.L.	N.L.
Precision	$\frac{2 + 8}{20} = 50\%$	$\frac{8 + 5}{25} = 52\%$
Base recall	$\frac{2}{4} = 50\%$	$\frac{3}{4} = 75\%$
Matched recall	$\frac{2 + 8}{2+8+5} = 66\%$	$\frac{8 + 5}{2+8+5} = 87\%$

TIME

The comparative times for the two methods will be obtained from the print-out records of the searches. Apart from a straight comparison, these can also be related to other aspects, such as the number of relevant documents retrieved.

APPENDIX B

SEARCH OUTPUT RECORDS

Question 1-01

Controlled language

434944	434940	434348	432575	431384	431012	429936✓	429881
429436	429013	427992	427976	427754	427058	427032	426807
426520	426042	423223	422203 ✓	421943	421085	419581	419328
419187	417817	416227 ✓	416200	416199	415196✓	415190	414654
414229✓	414226	414201	413837	413259	412562	412371	412240
411739	411334	410859	410528	333449✓	333440	333420	333398
331175	330661	329899	328579	328531	328199	325615	325294
325130	324945	324523	323514	322626	322445	322222	320589
320586	320154	319768	319549	318596	314600	314565	312584
312302	311285	310973	310466				

Natural Language

434934	434933	434930	429984	429969	429936✓	429013	427982
423432	422237	422203✓	421169	420400	417282	417261	417240
416227✓	416204	415829	415196✓	414229✓	413268	413263	412538
412257	410835	410494	333449✓	333426	333087	332414	331524
331458	330859	328584	325530	324591	324096	323613	323600
322405	321441	313551	312945	312570	311706	311509	310560
310212							

Question 1-02

Controlled language

433665✓	432631✓	430107✓	421811 ✓	418335✓	417902 ✓	412882✓	410652
331477	326446	325572	324199	323571	323254 ✓	320993	320569
320200 ✓	318523	312748✓					

Natural Language

435128	434550	433665✓	432631✓	430107 ✓	427897	422533	422325
421811✓	421037	418335✓	417902 ✓	412882✓	327831	323254✓	323237
320200 ✓	312748✓						

Manual search

412882	421811	417902	432631

412882 421811
417902 432631

Question 1-03

Controlled language

429535	429529	427672 ✓	426728	426703 ✓	420896 ✓	418887 ✓	418842
417911 ✓	417909 ✓	414136 ✓	413881	412217	410617 ✓	333120	331859
327567 ✓	322386 ✓	322100 ✓	320686 ✓	314206 ✓	312495 ✓	311440 ✓	

Natural Language

434371	433677	431620	431520	427677	427675	427672 ✓	426728
426703 ✓	426308	426307	421764	421706	420896 ✓	419681	418887 ✓
418848	418842 ✓	418244	417911 ✓	417909 ✓	416325	414920	414882
414530	414254	414136 ✓	413928	413909	412869	412868	412841
411666	411657	411291	411252	411029	410910	410617 ✓	333120
331860	331145	331142	330211	328543	327567 ✓	325986	322386 ✓
322100 ✓	320686 ✓	319906	314210	314206 ✓	313680	312495 ✓	312191
311688	311440 ✓	311175	310658	310234			

Manual search

418842 314206 333120

Question 1-04

Controlled language

435113	434362 ✓	434160 ✓	433136	430595	430099	428099	425711
425674	422310	418830	417682	417387	416368 ✓	415845	412884 ✓
412354	411018 ✓	410821	410619	333930	333570	331784	330834
330646	329592	329089	328864	328863	327768	327574	327121 ✓
327104 ✓	325717	323718	323709	323699	321833 ✓	319650	313852 ✓
311883	310204	310203					

Natural Language

434362	434160	431689	431688	431687	431685	429525	429280
428675	427630	426744	423393	421609	419965	416904	416897
416893	416368	412884	412870	412869	412868	412867	412866
412865	412864	412330	411700	411018	330916	329597	327862
327121	327029	326662	321833	324091	318371	313852	311397
311306	311140	310201					

Manual Search

313144 327104 415857 416904 425674

Question 1-05

Controlled language

435278	434973	434959	434360	434347	434341	434340	434132
434074	433598	433511	433001	432030	431543	430951	430397
430010	430006	430005	429585	429436	429435	429289	429288
429274	429216	429016	428990	428422	428197	428186	427963
427535	427534	427369	426949	426368	426069	423132	423129
423063	422502	422498	422419	422237	422227	421162	421067
420539	420085	419700	419522	418580	418551	418196	418140
418132	417586	417579	417510	417400	417399	417397	417293
417292	417287	417286	416614	416268	416264	416246	415660
415364	414697	413600	413597	412936	412498	411318	410945
410854	410723	410533	410273	410271	410270	410265	333866
333497	333489	333487	333486	333483	333482	333480	333478
333471	333469	333467	333466	333465	333464	333459	333457
331993	330225	329964	329818	329530	328590	328497	328083
327809	327508	327505	327455	327450	325969	325916	325595
325509	324935	324688	324262	324235	323866	323726	323617
323616	322474	322415	322397	321830	321471	321076	320942
320940	320608	320043	320021	319953	318965	318605	318594
318571	318144	314613	313928	313575	313331	312928	312693
312588	312235	312229	310961	310893	310483		

Natural Language

413600 413597

Manual search

410723 413597 417579 420539 333487

Controlled language

417510

Natural Language

435291	435144	433236	433233	433169	432605	430247	429675
429177	428945	428631	428167	427704	427390	426337	426195
423337	423350	422960	421400	421048	420698	420067	417513
417497	417495	416804	416498	414671	414453	414166	413503
413495	413193	412403	412279	411596	411531	410973	410945
410728	333682	332617	332616	332579	332119	331865	331687
331040	330843	330697	330696	329745	328944	328693	326746
325983	325965	325913	325840	325839	325760	325486	324719
323856	323797	323753	322719	322677	320941	320659	319823
319791	319701	319488	318735	318024	315101	314806	314794
314751	314746	314077	313782	313679	312844	311973	311751
310743	310703						

Manual search

417510 414962 411531 332738
330843

Question 107

Controlled language

434458 422510 ✓ 420544 ✓ 416978 ✓ 410242 ✓ 333850 ✓ 332782 331785 ✓
329962 ✓ 324943 324676 320937 318043 ✓ 314853 ✓ 311897
✗

Natural Language

434337 433307 422639 422510 ✓ 422508 422503 421898 421522
420544 ✗ 416978 ✓ 410825 410242 ✓ 333850 ✓ 331785 ✓ 331705 330250
329962 ✗ 326992 318846 318043 ✓ 313288 313021 312493 314853 ✓
311383

Manual search

416978 422510 314853

Question 1-08

Controlled language

435263 413546 431288 314440

// Natural Language \

435263 431288 430275 427350 426275 422488 421421 415487
413546 411628 411626 411624 332659 327746 320847 314833
314440 310816

/

Manual search

411626 413546 421421 435263 314440
430275 314833 332659

Question -1-09

Controlled language

425946 ✓ 417411 416112 ✓ 415518 ✓ 411201 328494 ✓ 327839 ✓
322394 ✓ 318452 ✓ 318451 ✓ 318450 ✓ 318449 ✓ 314946 ✓ 311386 ✓

433335 325484

Natural Language

433345 433335 ✓ 433334 428311 427343 426405 425946 ✓
417408 416112 ✓ 415518 ✓ 413695 413565 332644 329870
327840 327839 ✓ 327838 325473 322394 ✓ 320886 318452 ✓
318450 ✓ 318449 ✓ 314946 ✓ 310823 310482 311386 ✓

433346 417412 328494 318451

Manual search

433335 416112 417406 417412 433945 431907

Question 1-10

Controlled language

435287 434496 434389 433377 432358 432295 430329 429833
429796 429358 428412 423520 423517 417606 416597 414624
414549 414543 414541 414537 410839 333878 332795 331823
328890 326616 318930

Natural Language

434374 430329 ✓ 428412 ✓ 426365 423517 ✓ 423411 422516 414608
412539 332793 332484 331950 331802 328977 328890 ✓ 322414
318927 310983

Manual search

423517 428333 429538 422295

Question 2-01

Controlled Language

434952	434892	433989	433490	432696	431948	431720	429871
429860	429854	429833	429830	429829	429820	429794	429627
429620	429485	429262	428550	427922	427920	427689	427515
426776	426493	422849	422697	422692	421892	421863	421076
420080	419715	419376	418695	417888	417765	417193	416779
416484	415102	414780	414138	413878	413732	412849	411066
411059	410834	410747	410495	410469	333101	331096	329188
328126	327554	327039	326426	326086	325188	324535	324289
323521	323254	323237	323007	322736	322691	321621	321069
321021	321012	320799	320554	319670	319498	318024	315065
314168	313741	312706	312522	312471	311221	310272	310026

Natural Language

429636	429620	428776	428550	422692	420077	415125	410491
410474	410469	331456	331429	327471	324287	323562	321069
311280							

Question 2-02

Controlled Language

414236 429440

Natural Language

429440 421699 414236 313517

Manual Search

329583 331865 430007

Question 2-03

Controlled Language

433349 432050 427137 426867 426122 411382 323074 321511

Natural Language

435177	434014	433913	433865	433628	433551	433349	432846
432562	432252	432050	431116	429960	429427	429263	428979
428591	427928	427137	427132	427003	426867	426608	426580
426122	423368	422744	422604	422467	422445	422187	420717
420551	420205	419739	419262	419095	418095	417831	416305
415098	414502	414317	414316	413203	412499	412178	411880
411384	411383	411382	410815	410368	333840	332632	332508
332404	332297	331587	331369	331041	331013	330615	330382
330377	327807	327563	327314	327072	324644	324340	324339
323465	323099	323074	322467	322047	321700	321511	320847
320431	320340	319094	319089	318511	318381	315108	315093
314419	313970	313895	313386	311465	311395	311116	310542
310469	310437						

Manual Search

432050 426867 321511 414311 431116 315108 330615

Question 2-04

Controlled Language

427474 331261 325312 324332 322253 322250 311272 310514

Natural Language

431778	431774	429945	427756	422939	421951	419923	417993
417989	415991	410338	410019	331261	327226	327222	327221
326303	325312	324332	322253	311272			

Manual Search

32603 427756

Question 2-05

Controlled Language

434337	434334	434321	434314	433466	433441	433437	433430
433325	433319	432320	431760	431736	431488	431485	430638
430417	429277	427722	427385	426444	422628	422505	421653
421634	421523	421520	420553	419688	419525	418936	418679
418668	418663	418503	417709	417006	416251	415532	414767
414731	414714	414711	413674	413586	412714	412516	412511
412503	411821	410804	410801	410322	410028	410019	33297
332927	332926	331940	331746	331245	330936	330929	330928
330050	330021	328123	327924	326997	326880	326284	326281
326280	326000	325025	325024	324317	324281	324080	324058
324004	323351	322957	322900	322202	321830	320999	320946
318994	318008	318007	314869	314005	312964	311890	310317

Natural Language

434462	434321	434320	433466	433430	433325	433319	432409
432324	431736	431488	431485	429399	429394	429287	429277
426447	426444	425577	422628	421658	421653	421634	421520
420648	420529	419688	418663	418510	417709	417570	417504
417021	417006	416251	415532	414731	414714	413721	413719
413712	413674	412714	412707	412702	412516	411826	411821
411708	411707	410801	410028	410019	332974	332973	332954
332945	332927	332926	331940	330936	330929	330050	330021
328123	327924	327889	326997	326880	326280	326000	325291
325276	325025	325024	324317	324281	324080	324058	32333+
322957	322955	322202	320999	319998	318997	318995	318994
318000	314005	313278	312923	311890	311016	310886	310317
310030	310008						

Manual Search

318015 320995

Question 2-06

Controlled Language

435108	435105	435096	435095	434354	432099	431927	430902
428096	427887	427885	427883	423211	423210	423206	423204
423203	418120	417390	416116	415104	414352	412353	333566
331609	330656	329704	329428	327131	326661	326660	326658
324652	323723	320716	320712	320711	320709	320708	320700
320699	320697	320687	320685	320682	319649	319635	319634
314435	314432	312685	311689	311680			

Natural Language

435108	435096	435095	432100	432099	431927	431908	428960
428096	427887	427885	420230	419282	417388	415094	414354
413577	413142	412353	411080	410617	333566	333365	332788
331609	330853	330656	330645	327821	327567	324652	323895
322622	320709	320708	320703	320697	320688	320685	319636
313436							

Manual Search

312685 315685 315667

Question 2-07

Controlled Language

435106 ✓ 429228 428693 427676 427647 ✓ 426258 422131 421674
420568 420344 417923 415902 ✓ 415742 415516 414532 ✓ 414358
413888 412652 412641 ✓ 412640 412639 ✓ 412615 412598 ✓ 412597
411295 410882 410467 410054 331139 328907 326205 320724
320723 319231 311150 310733

Natural Language

435106 ✓ 434633 432633 431658 427721 427647 ✓ 426729 422849
421813 420234 417185 415902 ✓ 415886 414532 ✓ 413126 412642
412641 ✓ 412639 ✓ 412598 ✓ 412597 ✓ 328048 323733 314189 314176
312459 311193 311171

Manual Search

414358 415902 412639 412640 412641

Question 2-08

Controlled Language

Question 2-08

Controlled Language

434294 ✓ 434293 434283 ✓ 433291 ✓ 427208 419455 ✓ 311850 311848
311847 311841

Natural Language

435245 434294 ✓ 434283 ✓ 433291 ✓ 432274 429241 428057 427110
422947 422442 419473 419472 419470 419462 419457 419455
419454 418493 416032 410778 332662 331899 331898 331891
330843 311880 311854 311851 311846 311845 311844 311839
311837 311836 311813

Manual Search

311846 332662 434294

Question 2-09

Controlled Language

434349	432255	431880	430812	430543	430534	430304	429612
428675	426672	426324	423401	422970	422122	422121	422120
422119	422024	421900	420556	419508	418518	417581	417575
416893	414891	414890	414621	413594	412964	412933	412877
412870	412864	412512	412117	412107	332755	332098	331859
331470	331312	331211	330836	329878	328132	327853	327848
327521	327183	327100	325419	324985	324979	324640	324178
323904	323890	323698	323252	323109	322822	322820	321423
320889	320686	320174	319345	318860	318178	313335	313276
311909	311397	310287	310279				

Natural Language

434600	434354	434349	432308	432318	430812	430787	430543
430236	429612	427797	427361	425851	425702	423401	422970
422616	422475	422122	422121	422119	422024	421900	420556
420532	419522	419508	418516	417784	417581	417537	417293
416567	415547	415546	415463	415073	413594	413577	412964
412933	412877	412870	412864	412492	412354	412117	412107
411943	411197	410807	333238	332098	330816	328875	328132
327853	327848	327183	326662	325883	325039	324985	324979
323904	323876	323858	323532	323318	323252	321833	320686
320174	319888	319886	319345	318860	318086	313826	313646
313335	313276	312928	312904	312397	311909	311397	311306

Manual Search

Question 2-10

Controlled Language

427487	422654	422646	422644	320303	320006

Natural Language

433148	429385	429142	427487	426448	422663	422654	422646
422644	417756	417405	417004	413946	333748	333742	333612
332624	331699	330937	330241	330024	327923	327704	327528
325073	323735	322987	321573	320303	320006	319998	313031
312862	312036	312016	312015	312013	311800		

Manual Search

422646 417004 326687 327897

Question 4.01

Controlled Language

433637 ✓	432628	432613 ✓	429521 ✓	429515 ✓	427631 ✓	426752	425718
420831	419301	417894	417866	415862 ✓	414914 ✓	412912	412855
411024	331121	331120 ✓	331070	331069 ✓	331068 ✓	331067 ✓	330644
329094 ✓	327128 ✓	325191	324658	324191 ✓	323529	323269 ✓	323265 ✓
322103	322072	320206 ✓	320195 ✓	319197	313222 ✓	313094	311149 ✓
310248	310227						

Natural Language

434211	433637 ✓	433118	432860	432613 ✓	431679	431294	429521 ✓
429515 ✓	428672	427631 ✓	425996	417220	416368	416184	415862 ✓
415853	414914 ✓	411963	411077	332076	331120 ✓	331069 ✓	331068 ✓
331067 ✓	331066	331065	331055	330062	329094 ✓	327128 ✓	325656
325541	325299	324191 ✓	323898	323886	323568	323269 ✓	323265 ✓
322172	321825	320681	320206 ✓	320195 ✓	319969	314523	313502
313411	313222 ✓	311149 ✓	310832	310523			

Manual Search

320206 322072 323625 324191 331120

Question 4.02

Controlled Language

433152 ✓ 432866 430879 429969 326429 323486 ✓

435042 434300 423387 421977 418085 331374 329222 325741

312437 ✓

Natural Language

434719	434559	433152 ✓	433137	433093	429704	426132	426047
425954	425894	422998	422810	422344	419442	418033	416302
416069	414872	414796	414268	414009	412984	412455	412160
411129	411099 ✓	410442	333176	333159	331225	327937	327329
324539	323486 ✓	322743	322628 ✓	321523	321388	321356	318839
318385	315822	312693	312437 ✓				

Manual Search

312437 315716 321356 322381

Question 4.03

Controlled Langauge

430034	410579	329880	329879	328281	326320	329394	328458
314343	313364	433841	433840	433347	430036	429724	429051
429045	333790	329615	322575	318658	313626	310616	430041
430034	410579	329880	329879	328281	326320		

Natural Language

432083	426581	330857	330420

Manual Search

316394	313626	316402	316403

Question 4.04

Controlled Language

453837	429734	320456

Natural Language

422434	422433	415413	412454	326815	321749

Manual Search

-

Question 4-05

Controlled Language

410747

Natural Language

430919	426806	423344	423343	423278	418134	415461	414372
412444	410713	333747	332179	332030	330483	323855	321408
320299	319458	313471	311280				

Question 4-06

Controlled Language

434635	434425	433680	433679	433650	432609	422970	419850
418850	418849	418846	414619	327816	327521	325237	325183
323252	318214	313207	310207	310206	310187		

433721

373	433720	426764	413933	413886	332098	331112	326193

310208 310244 320288

Question 4.07

Controlled Language

433063 430263 429269 427835 427126 426277 422430 422428
416515 413447 411617 333778 332656 331872 331756 331711
330812 330770 327736 326614 314065

Natural Language

435215 429269 426277 426273 422433 422431 422428 421466
421412 420458 415475 414498 331877 331872 331711 328798
323701 321747 321706

Manual Search

331711 422428 430263

Question 4.08

Controlled Language

430959 428264 427345 423289 421094 419447 419429 412393
412276 329823 324440 318716 311734 311113

Natural Language

434193 430513 428661 423289 421094 421093 421016 419694
419448 419212 419169 419154 419875 416527 416566 415090
412276 332654 332537 332146 331722 329823 326702 325150
324440 321422 321417 321399 318716 314559 311734 311549

Manual Search

324440
318716
419694

Question 4.09

Controlled Language

432893	432822	432788	430728	430714	430692	429727	428892
427799	427798	426876	422050	422049	422046	422024	420964
420953	418976	418040	417062	414076	414035	412117	411155
410361	332266	331359	331357	331331	330317	329276	328452
328309	328223	328214	328213	328212	328211	328210	327268
326322	325910	325391	323453	322336	322287	321361	319342
314367	314323	313335	313070	312404	312403	310358	

Natural Language

432810	432809	432788	431861	431788	430720	430718	430714
430713	430712	430711	430710	430709	430708	430692	429727
428938	428892	428867	428830	428829	427849	427800	427798
425857	422770	422024	422023	421998	420964	420953	420001
419032	418976	418971	418961	418040	417099	417074	417059
415034	415019	414076	414035	413071	412134	412121	412117
411165	411155	410361	333309	333251	332269	332257	331357
331319	331289	331288	330288	329276	328406	328396	328309
328307	328275	328237	328232	328231	328230	328229	328228
328226	328222	328214	328213	328211	328210	328209	327357
327280	326350	326340	326322	325910	325732	325365	324399
322288	322287	321361	321334	320352	320348	319367	319342
318376	318359	318355	314367	314324	314323	313335	313333
313070	312331	311325	310358				

Manual Search

423810 331359 313335 427797

Question 4.10

Controlled Language

433560	432519	426612	426573	422748	420754	416534	331002
329029	326069	325137	325108	322041	314105	311053	310091
310086	310076						

Natural Language

433560	432540	432539	432519	429456	428646	428623	426573
422755	422748	421752	420754	420753	420732	417808	416861
416828	416534	333047	331002	329029	327959	326069	325137
325108	322041	322000	321984	321983	321981	318117	313457
313100	311053	310091	310086	310076			

Manual Search

311086 315126 316054 331002

Question 5.01

Controlled Language

432208 430559 420098 417482 416453 415432 411579 333706
329770 326790 321643 320795 320792 319059

Natural Language

427270 421376 421356 413470 410496 333706 323783 320795

Manual Search

429196 411055 426215 418374

Question 5.02

Controlled Language

434361 429293 330440 329888 320904 320725 320717 320711
320686 320684 318953 313047

Natural Language

432100 429293 427926 426152 422477 417388 415387 414354
414347 410617 333566 332788 331549 330440 327567 324652
323895 320717 320711 320709 320705 320688 320686 320685
313776 311689

430902 427175 423208 330645 329429 326658 320708 319630
314439 314438 312685

Manual Search

320717 320684 320705 320685

Question 5.03

Controlled Language

434769 431930 422026 418122 326153 323514 312439

Natural Language

431930 431715 431709 422956 420536 417167 413090 411290
411204 327397 320390 320353 313457 312234 311490 311296

Manual Search

330269

Question 5.04

Controlled Language

429431

Natural Language

429431 413768

Manual Search

429431

Question 5.05

434549 433512 432470 431536 430221 429428 426508 422705
422704 421685 419698 414788 323976 319981 313050 312062

Natural Language

434360 432900 432331 431953 429427 428544 427538 416899
414787 413758 410947 331995 331993 330982 330470 329011
326487 325913 325096 324234 321959 320040 318887 312061

434549 433512 432470 431536 430448 429428 426508 422705
422704 421685 419698 414788 330977 330470 323976 321973
319981 313160 313050 312062

Manual Search

412685 434549

Question 5.06

Controlled Language

433661 415904 320228 418884 421892 415898 430229
422259 416949 411991

Natural Language

430569 421891 413599 413598

Manual Search

421891 421892

QUESTION 5.07

Controlled Language

434652	434607	433767	433760	433672	433136	433127	432630
432150	431698	431696	431691	431683	431679	431678	431674
431638	431188	431186	431058	430592	429587	429586	428681
426663	425757	425704	420824	418838	417894	416948	416919
415918	415863	414910	414891	414343	413866	413860	413858
413857	412981	412880	411949	411943	411065	411016	410682
332676	332672	332540	332521	332150	327841	327728	327152
327137	327106	326166	326160	326159	325227	325200	325184
324250	324242	323264	322360	322169	322126	321233	321193
320211	314096	313174	313172	313167	312171	311210	310834
310213							

Natural Language

434652	434607	433760	432630	432150	430592	429586	426008
425784	425757	417894	416919	412881	412880	327841	326166
325200	324250	320211	313174	311210			

Manual Search

524250 413857 429586 320211 412880

Question 5.08

Controlled Language

430549	430530	430529	430524	425944	420810	420809	419806
415834	332155	331117	323271	318225	312143	311142	

Natural Language

432598	431636	430556	430530	430529	430524	427624	422834
422832	420810	420809	420234	417895	415869	412847	411230
411032	410511	410173	410132	332046	331117	331106	331098
331095	330131	330120	328084	327843	327039	326426	324656
324219	323326	323256	312195	314127	313187	313149	312199
312151	312143	311142	310269	310215	310196	310124	

Manual Search

430530

Question 5.09

Controlled Language

Question 5.09

Controlled Language

421210 415368 332748 330562 323136
427075 426758 426098 421194 330360 329601 325563 318937

434499 433007 429032 425554 421339 421200 420177 417920
416231 412993 412546 328675 327519 325654 325405 320639

429586 425784 418239 418087 416713 416278 415864 415307
413293 410567 329593 327524 326617 323850 323665 311468

Natural Language

434499 433007 432047 431511 430552 428053 427075 426758
426098 423218 421530 421339 421200 421194 420177 419909
419548 418548 417920 417315 416606 415927 414910 414608
414600 410830 333148 332484 330862 330562 330360 329581
329577 328977 328001 327609 327519 326617 325654 325405
324910 324823 319386 319001 318937 310588

Manual Search

417920 426098 427075

Question 4.10

Controlled Language

434537 433475 433433 432455 432435 432428 432417 4315
431492 430623 428782 421753 428529 421636 420677 4199
418643 417733 417731 416497 416377 415710 414978 4147
413713 412727 412718 41 815 410013 332976 332932 3329
331245 330914 330244 330059 329996 329179 326304 3262
326027 326017 325292 325265 324341 321938 321932 3219
321068 321037 321032 320335 318297 318289 315036 3140
312308 311229

435036 434716 434715 433951 433950 433816 430418 4284
428485 428484 422625 421528 418705 417767 327030 3252
322959 321080 321040 319269 312033

Natural Language

434716 434715 434707 434537 433951 433950 433475 43343
432725 432435 432428 431500 431492 430418 428782 42976
423757 428529 428499 428485 426791 422941 422630 42262
420639 419648 418906 417767 417733 417732 417731 41702
415992 415959 415957 414984 414978 414379 413713 41300
411723 411128 410342 333191 333169 332925 331245 33028
329179 329151 328201 328185 328136 326908 326288 32627
326027 325291 325265 325089 323342 322251 321938 32103
321314 320030 320003 318297 315036 314986 314958 31331
312033 311721 311229 310316

Manual Search

434715 433915 431500 422625 417731

434715 422625
43395 417731
431500

APPENDIX C

MASTER RECORD SHEETS

QUESTION 1-01

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	Relevance 2	Relevance 3	
1	434944	C		✓		
2	431012	C			✓	
3	427992	C			✓	
4	426807	C			✓	
5	421943	C			✓	
6	419187	C			✓	
7	415190	C			✓	
8	413837	C			✓	
9	411739	C			✓	
10	333440	C		✓		
11	329889	C			✓	
12	325294	C			✓	
13	322626	C			✓	
14	320154	C			✓	
15	314565	C		✓		
16	423934	N			✓	
17	429936	N			✓	
18	422203	N		✓		C
19	417240	N			✓	
20	414222	N			✓	
21	412257	N			✓	
22	333087	N			✓	
23	328584	N			✓	
24	323600	N			✓	
25	312570	N			✓	T

ANALYSIS

	C'L.	N.L.
Items Retrieved	70	49
Overlap	6	
Search time	37	28
Pre-search time	30	15
Total sets	28	17
Search sets	18	13
Combine sets	10	4

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	0	0	0
Relevance 2	3	1	0
Non-relevant	12	9	1
Ratios			
Relevance 1 & 2			
Precision	20	10	
Base Recall	—	—	
Matched Recall	100	25	
Relevance 1			
Precision	0	0	
Base Recall	—	—	
Matched Recall	—	—	

RELEVANT BASE DOCUMENTS

Relevance 1 —

Relevance 2 —

SEARCHER MODE

A CL + NL

B NL + Concept file

QUESTION 1-02

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	435128	N		✓		
2	434550	N			✓	T
3	433665	CN		✓		
4	432631	CNM		✓		T
5	430107	CN			✓	
6	427897	N			✓	
7	422533	N			✓	
8	422325	N		✓		
9	421811	CNM		✓		
10	421037	N		✓		
11	418335	CN			✓	
12	417902	CNM	✓			T
13	412882	CNM	✓			T
14	327831	N			✓	
15	323254	CN			✓	
16	323237	N		✓		
17	320200	CN	✓			T
18	312748	CN	✓			T
19	331477	C			✓	
20	326446	C			✓	
21	325572	C			✓	
22	324199	C		✓		
23	323571	C			✓	
24	320993	C			✓	
25	320569	C			✓	

ANALYSIS

	C'L.	N.L.
Items Retrieved	17	15
Overlap	10	
Search time	37	47
Pre-search time	30	30
Total sets	17	15
Search sets	13	11
Combine sets	4	4

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	4	4	4
Relevance 2	4	7	1
Non-relevant	9	7	1
Ratios			
Relevance 1 & 2			
Precision	47	61	
Base Recall	100	100	
Matched Recall	66	91	
Relevance 1			
Precision	24	22	
Base Recall	100	100	
Matched Recall	100	100	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
417902	
412882	432631
	421811

SEARCHER MODE

A CL

B NL

QUESTION 1-03

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	429535	C		✓		
2	426703	CN		✓		
3	414136	C		✓		N
4	412217	C			✓	
5	327567	C		✓		N
6	314206	CNM		✓		
7	434371	N			✓	
8	427677	N	✓			
9	421706	N			✓	
10	418848	N		✓		
11	410325	N			✓	
12	414530	N		✓		
13	413909	N			✓	
14	331860	N			✓	
15	328543	N			✓	
16	322100	N		✓		C
17	311688	N			✓	
18	310234	N		✓		
19	411666	N			✓	
20	411029	N			✓	
21	418842	M		✓		CN
22	418887	C			✓	
23	333120	M	✓			CN
24	427672	C		✓		
25	431520	N			✓	

ANALYSIS

	C'L.	N.L.
Items Retrieved	23	61
Overlap	16	
Search time	17	53
Pre-search time	20	10
Total sets	15	30
Search sets	9	14
Combine sets	6	16

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	0	1	0
Relevance 2	6	6	0
Non-relevant	2	10	0
Ratios			
Relevance 1 & 2			
Precision	75	41	
Base Recall	100	100	
Matched Recall	69	93	
Relevance 1			
Precision	0	6	
Base Recall	100	100	
Matched Recall	50	100	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
333120	314206
	418842

SEARCHER MODE

A CL + NL.

B NL + concept file

QUESTION 1-04

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	Relevance 2	Relevance 3	
1	435103	C		✓		
2	430299	C		✓		
3	418530	C		✓		
4	412884	C	✓			N
5	333930	C			✓	
6	329592	C		✓		
7	327574	C	✓			
8	323709	C	✓			
9	311883	C			✓	
10	434160	C		✓		N
11	434362	N	✓			C T
12	431655	N		✓		
13	426744	N		✓		T
14	416597	N			✓	
15	410368	N		✓		C
16	411018	N	✓			C T
17	327121	N	✓			C T
18	321833	N		✓		C
19	313852	N	✓			C
20	310201	N			✓	
21	312144	M			✓	
22	327104	M		✓		CN
23	415857	M			✓	
24	416904	M	✓			N
25	425674	M	✓			C

ANALYSIS

	C'L.	N.L.
Items Retrieved	43	44
Overlap	9	
Search time	45	41
Pre-search time	30	30
Total sets	25	17
Search sets	4	12
Combine sets	21	12

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	3	4	3
Relevance 2	5	4	2
Non-relevant	2	2	0
Ratios			
Relevance 1 & 2			
Precision	80	80	
Base Recall	66	66	
Matched Recall	87	62	
Relevance 1			
Precision	30	40	
Base Recall	50	50	
Matched Recall	100	71	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
416904	327104
425674	

SEARCHER MODE

A CL

B NL

QUESTION 1-05

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	413600	N		✓		
2	413597	NM	✓			C
3	435278	C			✓	
4	434074	C			✓	
5	430210	C			✓	
6	429274	C			✓	
7	427535	C			✓	
8	423660	C			✓	
9	420539	CM			✓	
10	415130	C			✓	
11	415364	C			✓	
12	410854	C			✓	
13	333497	C			✓	
14	333471	C			✓	
15	331993	C			✓	
16	327809	C			✓	
17	325509	C			✓	
18	323616	C			✓	
19	320940	C			✓	
20	318571	C	✓			
21	312588	C			✓	
22	410723	M			✓	
23	417579	M		✓		C
24	333487	M			✓	
25	424132	C			✓	

ANALYSIS

	C'L.	N.L.
Items Retrieved	158	2
Overlap	2	
Search time	50	70
Pre-search time	30	10
Total sets	33	44
Search sets	23	34
Combine sets	10	10

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	1	1	0
Relevance 2	0	1	0
Non-relevant	19	0	0
Ratios			
Relevance 1 & 2			
Precision	5	100	
Base Recall	100	50	
Matched Recall	66	66	
Relevance 1			
Precision	5	50	
Base Recall	100	100	
Matched Recall	100	50	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
413597	417579

SEARCHER MODE

A CL + NL

B NL + Concept file

QUESTION 1-06

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	417510	CM	✓			
2	330843	NM			✓	
3	414962	M			✓	
4	332735	M			✓	
5	411531	M		✓		N
6	312844	N			✓	
7	314751	N			✓	
8	316735	N			✓	
9	—					
10	320659	N			✓	
11	323797	N			✓	
12	325839	N			✓	T
13	326746	N			✓	
14	330697	N		✓		
15	332119	N			✓	
16	410728	N			✓	
17	412279	N			✓	
18	414166	N			✓	
19	417495	N			✓	T
20	421648	N			✓	
21	426195	N			✓	
22	428651	N		✓		
23	430247	N			✓	
24	432605	N			✓	
25	435291	N			✓	

ANALYSIS

	C'L.	N.L.
Items Retrieved	1	90
Overlap	0	
Search time	49	40
Pre-search time	10	10
Total sets	22	33
Search sets	4	22
Combine sets	18	11

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	1	0	0
Relevance 2	0	2	0
Non-relevant	0	18	0
Ratios			
Relevance 1 & 2			
Precision	100	10	
Base Recall	50	50	
Matched Recall	25	75	
Relevance 1			
Precision	100	0	
Base Recall	100	0	
Matched Recall	100	0	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
417510	411531

SEARCHER MODE

A N.L.

B C.L.

QUESTION 1-07

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	434458	C			✓	
2	420544	CN	✓			T
3	416978	CM	✓			N
4	410242	CN	✓			T
5	332752	C			✓	
6	329962	CN	✓			T
7	324676	C		✓		
8	318403	C			✓	
9	311597	C		✓		
10	331755	C			✓	
11	434337	N			✓	
12	422639	N			✓	
13	422510	NM	✓			C T
14	422508	N			✓	
15	421898	N			✓	
16	421522	N			✓	
17	410825	N			✓	
18						
19	333850	N	✓			C T
20	331705	N		✓		
21	330250	N			✓	
22	318846	N		✓		
23	313288	N			✓	
24	314853	NM		✓		C T
25						

ANALYSIS

	C'L.	N.L.
Items Retrieved	15	25
Overlap	9	
Search time	54	37
Pre-search time		45
Total sets	34	36
Search sets	25	25
Combine sets	9	11

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	4	5	5
Relevance 2	2	3	1
Non-relevant	4	8	0
Ratios			
Relevance 1 & 2			
Precision	60	53	
Base Recall	100	100	
Matched Recall	82	82	
Relevance 1			
Precision	40	31	
Base Recall	100	100	
Matched Recall	100	100	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
416978	314853
422510	

SEARCHER MODE

A NL + Concept

B CL + NL

QUESTION 1-08

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	432563	CNM	✓			T
2	431258	CN	✓			T
3	413546	CNM			✓	T
4	314440	CNM		✓		T
5	430275	NM	✓			T
6	427350	N			✓	T
7	426275	N			✓	
8	422455	N			✓	T
9	421421	NM			✓	
10	415487	N			✓	
11	411625	N			✓	T
12	411626	NM		✓		T
13	411624	N		✓		T
14	332659	NM		✓		
15	327746	N			✓	T
16	320847	N			✓	
17	314833	NM	✓			T
18	310516	N			✓	
19						
20						
21						
22						
23						
24						
25						

ANALYSIS

	C'L.	N.L.
Items Retrieved	4	18
Overlap	4	
Search time	20	12
Pre-search time	35	35
Total sets	12	6
Search sets	2	1
Combine sets	12	6

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	2	4	4
Relevance 2	1	4	3
Non-relevant	1	10	5
Ratios			
Relevance 1 & 2			
Precision	75	44	
Base Recall	28	100	
Matched Recall	37	100	
Relevance 1			
Precision	50	22	
Base Recall	33	100	
Matched Recall	50	100	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
432563	314440
430275	332659
314833	411626
	411624

SEARCHER MODE

A N.L.

B C.L.

QUESTION 1-09

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	433335	CNM	✓			
2	417411	C			✓	
3	416112	CNM		✓		T
4	410201	C			✓	
5	328494	C		✓		N
6	327839	C		✓		N
7	318452	C		✓		N
8	318450	CN		✓		T
9	311386	CN			✓	
10	433346	N			✓	T
11	433344	N		✓		
12	427343	N	✓			
13	425946	N		✓		C
14	413695	N			✓	
15	329570	N			✓	
16	327840	N			✓	T
17	325473	N			✓	
18	320886	N			✓	
19	314946	N			✓	T
20	322394	C		✓		N
21	413565	N			✓	
22	417406	M			✓	
23	417412	MN			✓	
24	433945	M			✓	
25	431907	M			✓	

ANALYSIS

	C'L.	N.L.
Items Retrieved	16	30
Overlap	13	
Search time	24	15
Pre-search time	30	0
Total sets	30	20
Search sets	25	17
Combine sets	5	3

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	1	2	0
Relevance 2	6	4	2
Non-relevant	3	10	3
Ratios			
Relevance 1 & 2			
Precision	70	38	
Base Recall	100	100	
Matched Recall	80	100	
Relevance 1			
Precision	10	12	
Base Recall	100	100	
Matched Recall	50	100	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
433335	416112

SEARCHER MODE

A NL + concept

B CL + NL

QUESTION 1-10

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	434374	N		✓		
2	435287	C	✓			
3	434359	C	✓			
4	432358	C		✓		
5	430329	CN	✓			
6	429746	C		✓		
7	428412	C	✓			N
8	422516	N	✓			
9	432295	M	✓			C
10	428333	M	✓			
11	429358	M	✓			C
12	423517	MN		✓		C
13	416597	C		✓		
14	414549	C		✓		
15	414541	C		✓		
16	412539	N		✓		
17	410539	C	✓			
18	332795	C		✓		
19	332484	N	✓			T
20	331502	N	✓			T
21	328590	C	✓			N
22	315930	C		✓		
23	315927	N		✓		T
24	434496	C		✓		
25	426365	N		✓		

ANALYSIS

	C'L.	N.L.
Items Retrieved	27	15
Overlap	4	
Search time	23	22
Pre-search time	–	0
Total sets	16	15
Search sets	13	11
Combine sets	3	4

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	6	4	2
Relevance 2	8	5	1
Non-relevant	0	0	0
Ratios			
Relevance 1 & 2			
Precision	100	100	
Base Recall	75	25	
Matched Recall	68	50	
Relevance 1			
Precision	43	44	
Base Recall	66	0	
Matched Recall	66	66	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
432295	423517
428333	
429358	

SEARCHER MODE

A N L.

B C. L.

QUESTION 2-01

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	434952	C	✓			
2	434892	C			✓	
3	429871	C			✓	
4	429830	C			✓	
5	429485	C			✓	
6	427515	C		✓		
7	421892	C			✓	
8	418695	C		✓		
9	415142	C		✓		
10	411066	C			✓	
11	431720	C			✓	
12	429829	C		✓		
13	426776	C			✓	
14	321069	CN	✓			
15	419376	C			✓	
16	416484	C			✓	
17	429620	N		✓		C
18	428550	N			✓	
19	420077	N		✓		
20	410491	N		✓		
21	410469	N			✓	
22	331429	N			✓	
23	324287	N			✓	
24	323562	N		✓		
25	311280	N		✓		

ANALYSIS

	C'L.	N.L.
Items Retrieved	85	16
Overlap	4	
Search time	29	55
Pre-search time	5	0
Total sets	40	30
Search sets	31	21
Combine sets	9	9

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	2	1	0
Relevance 2	4	5	0
Non-relevant	10	4	0
Ratios			
Relevance 1 & 2			
Precision	38	60	
Base Recall	—	—	
Matched Recall	63	54	
Relevance 1			
Precision	13	10	
Base Recall	—	—	
Matched Recall	100	50	

RELEVANT BASE DOCUMENTS

Relevance 1 Relevance 2

SEARCHER MODE

A CL + NL

B NL + Concept File

QUESTION 2-02

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	329533	M			✓	
2	331865	M			✓	
3	313517	N			✓	
4	421699	N			✓	
5	414236	NC		✓		
6	429440	NC		✓		
7	430007	NM		✓		
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						

ANALYSIS

	C'L.	N.L.
Items Retrieved	2	5
Overlap	2	
Search time	7	39
Pre-search time	5	5
Total sets	6	12
Search sets	4	8
Combine sets	2	4

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	0	0	0
Relevance 2	2	3	0
Non-relevant	0	2	0
Ratios			
Relevance 1 & 2			
Precision	100	60	
Base Recall	0	100	
Matched Recall	67	100	
Relevance 1			
Precision	0	0	
Base Recall	—	—	
Matched Recall	—	—	

RELEVANT BASE DOCUMENTS

Relevance 1 Relevance 2

430007

SEARCHER MODE

A C.L.

B N.L.

QUESTION 2-03

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	423349	CN		✓		T
2	432050	CM	✓			N T
3	427137	CN	✓			
4	426867	CNM	✓			
5	426122	CN		✓		
6	411382	C	✓			N T
7	323074	C	✓			N T
8	321511	CNM	✓			T
9	435177	N			✓	
10	429960	N			✓	
11	422187	N			✓	
12	419095	N		✓		
13	414317	N	✓			
14	411384	N	✓			
15	433865	N			✓	
16	432252	N		✓		
17	428979	N			✓	
18	422604	N	✓			
19	420205	N		✓		
20	416305	N	✓			
21	412499	N			✓	
22	315105	M			✓	
23	431116	M	✓			N
24	414311	M		✓		
25	330615	M	✓			N

ANALYSIS

	C'L.	N.L.
Items Retrieved	8	98
Overlap	8	
Search time	35	65
Pre-search time	5	0
Total sets	29	43
Search sets	20	30
Combine sets	9	13

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	6	7	4
Relevance 2	2	5	1
Non-relevant	0	6	0
Ratios			
Relevance 1 & 2			
Precision	100	67	
Base Recall	50	83	
Matched Recall	54	100	
Relevance 1			
Precision	75	39	
Base Recall	60	100	
Matched Recall	60	100	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
330615	414311
431116	
321511	
426867	
432050	

SEARCHER MODE

A C.L. + N.L.

B N.L. + concept

QUESTION 2-04

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	427474	C			✓	
2	331621	CN			✓	
3	325312	CN	✓			
4	324332	CN		✓		
5	322253	CN	✓			T
6	322250	C			✓	
7	311272	CN			✓	
P	310514	C			✓	
9	431778	N		✓		
10	431774	N			✓	
11	429945	N			✓	
12	427756	NM		✓		
13	422939	N			✓	
14	421951	N			✓	
15	419923	N		✓		
16	417993	N			✓	
17	417989	N			✓	
18	410338	N	✓			
19	415991	N		✓		
20	410019	N			✓	
21	327226	N		✓		
22	327222	N			✓	
23	327221	N		✓		
24	326303	NM		✓		
25						

ANALYSIS

	C'L.	N.L.
Items Retrieved	8	21
Overlap	5	
Search time	25	44
Pre-search time	10	0
Total sets	29	26
Search sets	26	16
Combine sets	3	8

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	2	3	1
Relevance 2	1	8	0
Non-relevant	5	10	0
Ratios			
Relevance 1 & 2			
Precision	37	52	
Base Recall	0	100	
Matched Recall	27	100	
Relevance 1			
Precision	25	14	
Base Recall	-	-	
Matched Recall	67	100	

RELEVANT BASE DOCUMENTS

Relevance 1 Relevance 2

326303
427756

SEARCHER MODE

A C.L.

B N.L.

QUESTION 2-05

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	434337	C			✓	
2	433325	C			✓	I
3	430417	C			✓	
4	421634	C		✓		N
5	418668	C		✓		N
6	414731	C			✓	
7	412503	C			✓	
8	332927	C			✓	
9	330050	C		✓		N
10	326280	C	✓			N
11	326004	C		✓		
12	318904	C	✓			N
13	434462	N			✓	
14	432324	N			✓	
15	426447	N	✓			
16	420648	N		✓		
17	417021	N			✓	
18	415532	N			✓	
19	413712	N			✓	
20	412707	N	✓			
21	411708	N			✓	
22	411707	N			✓	
23	318015	M		✓		
24	320995	M	✓			
25	411821	N		✓		C

ANALYSIS

	C'L.	N.L.
Items Retrieved	96	98
Overlap	56	
Search time	42	41
Pre-search time	10	0
Total sets	39	29
Search sets	28	18
Combine sets	11	11

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	2	2	0
Relevance 2	4	2	0
Non-relevant	6	7	0
Ratios			
Relevance 1 & 2			
Precision	50	36	
Base Recall	0	0	
Matched Recall	70	90	
Relevance 1			
Precision	17	18	
Base Recall	0	0	
Matched Recall	50	100	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
320995	318015

SEARCHER MODE

A C.L.

B N.L.

QUESTION 2-06

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	Relevance 2	Relevance 3	
1	435105	CN			✓	
2	434354	C		✓		
3	425096	CN	✓			
4	423211	C		✓		
5	423203	C	✓			
6	415104	C	✓			
7	331609	CN	✓			
8	327131	C	✓			
9	324652	C	✓			N
10	320711	C		✓		
11	320699	C		✓		
12	320652	C	✓			
13	314435	C	✓			
14	314432	C	✓			
15	432049	N		✓		C
16	419252	N			✓	
17	413577	N			✓	
18	410617	N			✓	
19	327821	N		✓		
20	322622	N	✓			
21	320697	N	✓			C
22	313436	N		✓		
23	312655	M		✓		C
24	315655	M		✓		
25	315667	M	✓			

ANALYSIS

	C'L.	N.L.
Items Retrieved	53	41
Overlap	17	
Search time	30	45
Pre-search time	0	0
Total sets	30	40
Search sets	27	31
Combine sets	3	9

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	9	4	0
Relevance 2	4	3	0
Non-relevant	1	4	0
Ratios			
Relevance 1 & 2			
Precision	93	64	
Base Recall	33	0	
Matched Recall	82	44	
Relevance 1			
Precision	64	36	
Base Recall	0	0	
Matched Recall	90	45	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
315667	312655
	315655

SEARCHER MODE

A NL

B CL

QUESTION 2-07

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	435106	CN		✓		T
2	427671	C		✓		
3	422131	C	✓			
4	420344	C			✓	
5	415742	C			✓	
6	414358	CM	✓			
7	412641	CM	✓			N
8	412615	C		✓		
9	411295	C			✓	
10	410054	C	✓			
11	326205	C	✓			
12	319231	C	✓			
13	431658	N			✓	
14	426729	N		✓		
15	420234	N			✓	
16	415886	N			✓	
17	412642	N			✓	
18	412595	N	✓			C
19	323733	N			✓	
20	312458	N			✓	
21	415902	M		✓		CN
22	412639	M	✓			CN
23	412640	M	✓			C
24	427721	N			✓	
25	429225	C	✓			

ANALYSIS

	C'L.	N.L.
Items Retrieved	36	27
Overlap	8	
Search time	19	21
Pre-search time	10	0
Total sets	6	10
Search sets	5	6
Combine sets	1	4

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	7	1	0
Relevance 2	3	2	1
Non-relevant	3	7	0
Ratios			
Relevance 1 & 2			
Precision	77	30	
Base Recall	100	60	
Matched Recall	91	25	
Relevance 1			
Precision	54	10	
Base Recall	100	50	
Matched Recall	100	25	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
412639	415902
412640	
412641	
414358	

SEARCHER MODE

A NL + Concept

B CL + NL

QUESTION 2-08

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	434294	CM	✓			N
2	434253	NC		✓		
3	433291	C		✓		N
4	427205	C			✓	
5	311550	C		✓		
6	311847	C		✓		
7	435245	N			✓	
8	432274	N			✓	
9	425057	N			✓	
10	422947	N		✓		
11	419473	N		✓		
12	419470	N		✓		
13	419457	N		✓		
14	419454	N		✓		
15	416032	N			✓	
16	332662	NM	✓			T
17	331895	N		✓		T
18	330843	N			✓	
19	311854	N		✓		T
20	311846	NM	✓			T
21	311844	N			✓	T
22	311837	N	✓			
23	311813	N			✓	
24	434293	C		✓		
25	331899	N		✓		

ANALYSIS

	C'L.	N.L.
Items Retrieved	10	37
Overlap	4	
Search time	57	25
Pre-search time	0	0
Total sets	34	17
Search sets	25	12
Combine sets	9	5

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	1	3	2
Relevance 2	5	9	2
Non-relevant	1	7	1
Ratios			
Relevance 1 & 2			
Precision	85	63	
Base Recall	33	100	
Matched Recall	35	84	
Relevance 1			
Precision	14	16	
Base Recall	33	100	
Matched Recall	25	100	

RELEVANT BASE DOCUMENTS

Relevance 1 Relevance 2

434294
332662
311846

SEARCHER MODE

A NL

B CL

QUESTION 2-09

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	434349	C			✓	
2	429612	C			✓	
3	422121	C			✓	
4	418518	C			✓	
5	413594	C		✓		N
6	412117	C			✓	
7	331211	C		✓		
8	327183	C	✓			N
9	327521	C			✓	
10	432255	C			✓	
11	419505	C			✓	
12	434600	N			✓	
13	430543	N			✓	
14	423401	N			✓	
15	422024	N			✓	
16	422970	N			✓	
17	415463	N			✓	
18	412570	N		✓		C
19	411197	N			✓	
20	327853	N			✓	
21	324979	N		✓		C
22	321833	N			✓	
23	318080	N			✓	
24	422120	M			✓	
25	422517	M			✓	

ANALYSIS

	C'L.	N.L.
Items Retrieved	70	88
Overlap	35	
Search time	45	25
Pre-search time	0	0
Total sets	20	26
Search sets	14	16
Combine sets	6	10

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	1	0	0
Relevance 2	2	2	0
Non-relevant	8	10	0
Ratios			
Relevance 1 & 2			
Precision	18	16	
Base Recall	—	—	
Matched Recall	100	80	
Relevance 1			
Precision	9	0	
Base Recall	—	—	
Matched Recall	100	100	

RELEVANT BASE DOCUMENTS

Relevance 1 Relevance 2

SEARCHER MODE

A NL + concept

B CL + NL

QUESTION 2-10

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	427847	C	✓			N
2	422654	CN		✓		
3	422646	CM	✓			N
4	422644	CN	✓			
5	320303	CN	✓			T
6	320006	C	✓			N
7	433148	N		✓		T
8	429142	N		✓		T
9	426448	N			✓	
10	417405	N	✓			
11	413946	N	✓			T
12	333742	N		✓		
13	332624	N	✓			
14	330937	N		✓		
15	330024	N		✓		
16	327704	N		✓		
17	325073	N	✓			T
18	322987	N	✓			T
19	314998	N		✓		T
20	312862	N			✓	
21	312016	N	✓			
22	312013	N		✓		
23	417004	M	✓			N
24	326687	M	✓			
25	327897	M	✓			

ANALYSIS

	C'L.	N.L.
Items Retrieved	6	30
Overlap	6	
Search time	49	15
Pre-search time	0	0
Total sets	27	18
Search sets	22	13
Combine sets	5	5

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	5	8	4
Relevance 2	1	9	3
Non-relevant	0	2	0
Ratios			
Relevance 1 & 2			
Precision	100	90	
Base Recall	25	50	
Matched Recall	30	100	
Relevance 1			
Precision	83	42	
Base Recall	25	50	
Matched Recall	45	100	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
417004	
326687	
327897	
422646	

SEARCHER MODE

A NL

B CL

QUESTION 4-01

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	433637	C	✓			N
2	429521	C	✓			N
3	427631	C	✓			N
4	311149	C		✓		N
5	411024	C			✓	
6	324658	C		✓		
7	331067	C		✓		N
8	322072	CM		✓		N
9	415862	C		✓		N
10	434911	N		✓		T
11	432560	N		✓		T
12	431679	N		✓		
13	425996	N	✓			
14	415853	N		✓		
15	331069	N		✓		C T
16	331055	N		✓		
17	325299	N			✓	
18	323269	N		✓		C T
19	320195	N		✓		C T
20	313222	N		✓		C
21	—					
22	323265	M		✓		CN
23	324191	M		✓		CN
24	331120	M		✓		CN
25	419301	L		✓		

ANALYSIS

	C'L.	N.L.
Items Retrieved	42	53
Overlap	20	
Search time	30	40
Pre-search time	18	20
Total sets		28
Search sets		24
Combine sets		4

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	3	1	1
Relevance 2	6	9	5
Non-relevant	1	1	0
Ratios			
Relevance 1 & 2			
Precision	90	91	
Base Recall	100	100	
Matched Recall	68	90	
Relevance 1			
Precision	30	9	
Base Recall	—	—	
Matched Recall	75	100	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
	323625
	322072
	324191
	331120

SEARCHER MODE

A CL + NL

B NL + Concepts

QUESTION 4-02

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	433152	C		✓		N
2	430879	C		✓		
3	326429	C	✓			
4	435042	C			✓	
5	423387	C			✓	
6	418055	C			✓	
7	329222	C			✓	
8	434719	N			✓	
9	433137	N			✓	
10	426132	N			✓	
11	425894	N	✓			T
12	422344	N			✓	
13	416302	N			✓	
14	414796	N		✓		T
15	412984	N			✓	
16	411129	N			✓	
17	333176	N			✓	
18	327932	N		✓		T
19	323486	N		✓		C
20	321523	N			✓	
21	318839	N			✓	
22	315716	M			✓	
23	312437	M	✓			N
24	312356	M			✓	
25	322351	M			✓	

ANALYSIS

	C'L.	N.L.
Items Retrieved	15	44
Overlap	3	
Search time	30	40
Pre-search time	15	30
Total sets	18	25
Search sets	4	6
Combine sets	14	19

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	1	1	1
Relevance 2	2	3	2
Non-relevant	4	10	1
Ratios			
Relevance 1 & 2			
Precision	43	29	
Base Recall	0	100	
Matched Recall	57	71	
Relevance 1			
Precision	14	7	
Base Recall	0	100	
Matched Recall	50	50	

RELEVANT BASE DOCUMENTS

Relevance 1 Relevance 2

312437

SEARCHER MODE

A CL

B NL

QUESTION 4-03

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	Relevance 2	Relevance 3	
1	433842	C		✓		
2	433347	C			✓	
3	429051	C		✓		
4	333790	C			✓	
5	318658	C		✓		
6	430041	C			✓	
7	329850	C			✓	
8	326320	C		✓		
9	325458	C			✓	
10	313364	C			✓	
11	432083	N		✓		
12	430767	N		✓		
13	427120	N			✓	
14	416338	N			✓	
15	410521	N			✓	
16	332296	N		✓		T
17	328269	N		✓		
18	322568	N			✓	
19	320420	N	✓			T
20	314948	N	✓			
21	311662	N			✓	
22	316394	M	✓			
23	313626	M			✓	
24	316402	M		✓		
25	316403	M	✓			

ANALYSIS

	C'L.	N.L.
Items Retrieved	30	32
Overlap	0	
Search time	45	50
Pre-search time	12	30
Total sets	21	50
Search sets	16	34
Combine sets	5	16

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	0	2	1
Relevance 2	4	4	1
Non-relevant	6	5	0
Ratios			
Relevance 1 & 2			
Precision	40	54	
Base Recall	0	0	
Matched Recall	40	60	
Relevance 1			
Precision	0	18	
Base Recall	0	0	
Matched Recall	0	100	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
316394	316402
316403	

SEARCHER MODE

A CL + NL

B NL + Concepts

QUESTION 4-04

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	422434	N			✓	
2	422433	N			✓	
3	433837	C		✓		
4	429734	C			✓	
5	320456	C	✓			
6	415413	N		✓		
7	412454	N		✓		
8	326815	N		✓		
9	321749	N			✓	
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						

ANALYSIS

	C'L.	N.L.
Items Retrieved	3	6
Overlap	0	
Search time	38	30
Pre-search time	12	25
Total sets	32	33
Search sets	25	22
Combine sets	7	11

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	1	0	0
Relevance 2	1	3	0
Non-relevant	1	3	0
Ratios			
Relevance 1 & 2			
Precision	66	50	
Base Recall	–	–	
Matched Recall	40	60	
Relevance 1			
Precision	33	0	
Base Recall	–	–	
Matched Recall	100	0	

RELEVANT BASE DOCUMENTS

Relevance 1 —

Relevance 2 —

SEARCHER MODE

A CL

B NL

QUESTION 4-05

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	410747	C			✓	
2	430919	N		✓		T
3	426806	CNM		✓		T
4	423344	N			✓	
5	423343	N			✓	T
6	423278	N			✓	
7	415134	N			✓	
8	415461	N			✓	
9	414372	N			✓	
10	412444	N			✓	
11	410713	N			✓	
12	333747	N			✓	
13	332179	N			✓	
14	332000	N			✓	
15	330483	N			✓	
16	323855	N			✓	
17	321405	N			✓	
18	320299	N			✓	
19	319458	N			✓	
20	313471	N			✓	
21	311280	N			✓	
22						
23						
24						
25						

ANALYSIS

	C'L.	N.L.
Items Retrieved	2	20
Overlap		1
Search time	12	25
Pre-search time	10	15
Total sets	12	20
Search sets	10	11
Combine sets	2	9

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	0	0	0
Relevance 2	1	2	2
Non-relevant	1	18	1
Ratios			
Relevance 1 & 2			
Precision	50	10	
Base Recall	100	100	
Matched Recall	50	100	
Relevance 1			
Precision	0	0	
Base Recall	—	—	
Matched Recall	—	—	

RELEVANT BASE DOCUMENTS

Relevance 1 Relevance 2

426806

SEARCHER MODE

A CL + NL

B NL + Concept

QUESTION 4-06

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	434635	C		✓		
2	434425	C		✓		
3	433680	C			✓	
4	433650	C	✓			
5	432609	C			✓	
6	422970	C		✓		
7	418850	C		✓		
8	418546	C	✓			
9	418849	C			✓	
10	327516	C			✓	
11	327521	C		✓		
12	325207	C			✓	
13	323252	C		✓		
14	318214	C		✓		
15	313207	C			✓	
16	310206	C		✓		
17	433721	N	✓			
18	433720	N	✓			T
19	413933	N		✓		
20	413586	N	✓			
21	332098	N	✓			
22	326193	N	✓			
23	310201	M		✓		
24	310244	M	✓			
25	320288	M	✓			

ANALYSIS

	C'L.	N.L.
Items Retrieved	22	8
Overlap	0	
Search time	24	27
Pre-search time	20	20
Total sets	21	19
Search sets	5	7
Combine sets	16	12

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	4	5	1
Relevance 2	6	1	0
Non-relevant	6	0	0
Ratios			
Relevance 1 & 2			
Precision	62	100	
Base Recall	0	0	
Matched Recall	62	37	
Relevance 1			
Precision	25	83	
Base Recall	0	0	
Matched Recall	44	55	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
310244	310201
320288	

SEARCHER MODE

A NL

B CL

QUESTION 4-07

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	430263	CM			✓	
2	429629	CN	✓			T
3	427126	C			✓	
4	426277	C			✓	N
5	422425	CNM	✓			T
6	416515	C			✓	
7	411617	C		✓		
8	333775	C			✓	
9	331572	CN		✓		T
10	331756	C			✓	
11	330852	C	✓			
12	330770	C		✓		
13	326514	C			✓	
14	316065	C			✓	
15	435215	N	✓			T
16	426273	N		✓		T
17	422433	N		✓		T
18	421466	N	✓			T
19	420455	N			✓	T
20	415475	N		✓		T
21	331677	N			✓	T
22	325748	N	✓			T
23	328701	N	✓			T
24	321706	N		✓		T
25	331711	M	✓			CN

ANALYSIS

	C'L.	N.L.
Items Retrieved	21	19
Overlap	5	
Search time	20	15
Pre-search time	25	30
Total sets	20	18
Search sets	11	10
Combine sets	9	8

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	3	6	6
Relevance 2	3	5	3
Non-relevant	5	2	2
Ratios			
Relevance 1 & 2			
Precision	43	85	
Base Recall	100	100	
Matched Recall	43	75	
Relevance 1			
Precision	21	46	
Base Recall	100	100	
Matched Recall	43	55	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
422425	
331711	

SEARCHER MODE

A NL + Concept

B CL + NL

QUESTION 4-08

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	430959	C			✓	
2	428264	C			✓	
3	427345	C		✓		
4	421094	CN	✓			T
5	419447	C			✓	
6	418429	C			✓	
7	412276	CN			✓	
8	324440	CNM	✓			
9	311734	CN	✓			
10	434193	N	✓			T
11	430513	N			✓	
12	428661	N			✓	
13	421693	N			✓	
14	421016	N	✓			T
15	419645	N			✓	
16	419212	N			✓	
17	419169	N		✓		
18	418575	N			✓	
19	416466	N			✓	
20	332537	N			✓	
21	331722	N			✓	
22	326702	N			✓	
23	321417	N			✓	
24	318716	NM	✓			C
25	419644	M			✓	N

ANALYSIS

	C'L.	N.L.
Items Retrieved	14	32
Overlap	7	
Search time	32	42
Pre-search time	35	12
Total sets	47	32
Search sets	36	22
Combine sets	11	10

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	3	6	3
Relevance 2	1	1	0
Non-relevant	5	12	0
Ratios			
Relevance 1 & 2			
Precision	44	38	
Base Recall	100	100	
Matched Recall	63	88	
Relevance 1			
Precision	33	33	
Base Recall	100	100	
Matched Recall	67	100	

RELEVANT BASE DOCUMENTS

Relevance 1 Relevance 2

324440
318716

SEARCHER MODE

A NL

B CL

QUESTION 4-09

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	432893	C	✓			
2	425592	C		✓		N
3	422024	C		✓		N
4	414035	CN	✓			T
5	331331	C			✓	
6	325213	C		✓		N
7	325391	C			✓	
8	314323	C	✓			N
9	423510	NM	✓			
10	430714	N	✓			C
11	430713	N		✓		
12	430692	N			✓	C T
13	424727	N			✓	C
14	427849	N	✓			T
15	427800	N	✓			
16	430964	N		✓		C
17	421995	N	✓			T
18	415040	N		✓		C T
19	415961	N	✓			T
20	414076	N	✓			C T
21	411155	N		✓		C T
22	432758	N	✓			C
23	331359	M	✓			C
24	313335	M	✓			N
25	427797	M	✓			

ANALYSIS

	C'L.	N.L.
Items Retrieved	55	105
Overlap	25	
Search time	22	30
Pre-search time	25	12
Total sets	26	5
Search sets	20	5
Combine sets	6	3

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	3	9	5
Relevance 2	3	4	2
Non-relevant	2	2	1
Ratios			
Relevance 1 & 2			
Precision	75	87	
Base Recall	25	50	
Matched Recall	66	94	
Relevance 1			
Precision	35	60	
Base Recall	25	50	
Matched Recall	54	91	

RELEVANT BASE DOCUMENTS

Relevance 1 Relevance 2

427797
313335
331359
423510

SEARCHER MODE

A NL + Concept

B CL + NL.

QUESTION 4-10

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	433760	CN	✓			
2	432519	CN	✓			T
3	426612	N	✓			
4	426573	CN		✓		T
5						
6	416534	N		✓		C
7	331002	NM	✓			C
8	326069	N		✓		C
9	325105	N	✓			C
10						
11	310086	CNM	✓			T
12	432540	C		✓		
13	432539	C	✓			
14	429456	C			✓	
15	420754	C		✓		N
16	420732	C			✓	
17	416525	C		✓		
18	333047	C	✓			
19	327959	C			✓	
20	325137	C	✓			N
21						
22	321963	C			✓	
23	315117	C	✓			
24	315126	M	✓			
25	316054	M			✓	

ANALYSIS

	C'L.	N.L.
Items Retrieved	37	15
Overlap	12	
Search time	26	14
Pre-search time	15	10
Total sets	21	9
Search sets	16	7
Combine sets	5	2

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	7	6	2
Relevance 2	4	3	1
Non-relevant	4	0	0
Ratios			
Relevance 1 & 2			
Precision	74	100	
Base Recall	66	66	
Matched Recall	94	65	
Relevance 1			
Precision	48	67	
Base Recall	66	66	
Matched Recall	90	70	

RELEVANT BASE DOCUMENTS

Relevance 1 Relevance 2

315126
310086
331002

SEARCHER MODE

A NL
B CL

QUESTION 5-01

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	432205	C			✓	
2	430559	C		✓		
3	420095	C		✓		
4	417482	C	✓			
5	416453	C		✓		
6	415432	C		✓		
7	411579	C	✓			
8	333706	CN		✓		
9	329770	C		✓		
10	326790	C			✓	
11	321643	C			✓	
12	320795	CN			✓	
13	320792	C			✓	
14	319059	C			✓	
15	427270	N	✓			
16	421376	N			✓	
17	421356	N			✓	
18	413470	N			✓	
19	410490	N			✓	
20	323753	N		✓		
21	424190	M			✓	
22	411055	M			✓	
23	426315	M			✓	
24	415374	M			✓	
25						

ANALYSIS

	C'L.	N.L.
Items Retrieved	14	8
Overlap	2	
Search time	45	115
Pre-search time	10	15
Total sets	28	71
Search sets	21	44
Combine sets	7	27

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	2	1	0
Relevance 2	6	2	0
Non-relevant	6	5	0
Ratios			
Relevance 1 & 2			
Precision	57	35	
Base Recall	—	—	
Matched Recall	50	30	
Relevance 1			
Precision	14	12	
Base Recall	—	—	
Matched Recall	67	33	

RELEVANT BASE DOCUMENTS

Relevance 1 Relevance 2

SEARCHER MODE

A CL + NL

B NL + concept

QUESTION 5-02

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	434361	C			✓	
2	330440	C		✓		N
3	320904	C			✓	
4	320717	CM		✓		N
5	320688	C	✓			N
6	320684	CM	✓			
7	318953	C			✓	
8	432100	N	✓			
9	427926	N		✓		
10	422477	N			✓	
11	415387	N	✓			
12	414347	N		✓		T
13	333566	N		✓		T
14	351549	N			✓	
15	327567	N	✓			
16	328595	N			✓	
17	320711	N		✓		C T
18	320705	NM	✓			T
19	320685	NM	✓			T
20	430902	N		✓		
21	423208	N		✓		
22	329629	N		✓		
23	320708	N	✓			T
24	314439	N		✓		T
25	312685	N		✓		T

ANALYSIS

	C'L.	N.L.
Items Retrieved	12	37
Overlap	9	
Search time	35	74
Pre-search time	15	5
Total sets	31	51
Search sets	26	28
Combine sets	5	23
Descriptors		

PERFORMANCE

	C.L.	N.L.	
Totals			
Relevance 1	2	6	3
Relevance 2	2	9	5
Non-relevant	2	3	0
Ratios			
Relevance 1 & 2			
Precision	67	83	
Base Recall	50	75	
Matched Recall	21	95	
Relevance 1			
Precision	33	33	
Base Recall	33	66	
Matched Recall	25	87	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
320684	320717
320705	
320685	

SEARCHER MODE

A

B

QUESTION 5-03

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	434769	C			✓	
2	431930	CN			✓	
3	422026	C		✓		
4	418122	C	✓			
5	326153	C		✓		
6	323514	C		✓		
7	312439	C			✓	
8	431715	N			✓	
9	422956	N			✓	
10	420536	N			✓	
11	417167	N			✓	
12	413090	N		✓		
13	411290	N			✓	
14	411204	N		✓		
15	327397	N	✓			
16	320390	N			✓	
17	320353	N			✓	
18	313457	N	✓			
19	312234	N			✓	
20	311490	N		✓		
21	311296	N			✓	
22	431709	N			✓	
23						
24						
25						

ANALYSIS

	C'L.	N.L.
Items Retrieved	7	16
Overlap	1	
Search time	93	73
Pre-search time	10	10
Total sets	56	56
Search sets	40	39
Combine sets	16	17

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	1	2	0
Relevance 2	3	3	0
Non-relevant	3	11	0
Ratios			
Relevance 1 & 2			
Precision	57	32	
Base Recall	—	—	
Matched Recall	47	55	
Relevance 1			
Precision	14	12	
Base Recall	—	—	
Matched Recall	33	67	

RELEVANT BASE DOCUMENTS

Relevance 1 — Relevance 2 —

SEARCHER MODE

A CL + NL.

B NL + Concept

QUESTION 5-04

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	Relevance 2	Relevance 3	
1	429431	CNM	✓			
2	413768	N			✓	
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						

ANALYSIS

	C'L.	N.L.
Items Retrieved	1	2
Overlap	1	1
Search time	30	43
Pre-search time	15	0
Total sets	16	17
Search sets	11	13
Combine sets	5	4

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	1	1	1
Relevance 2	0	0	0
Non-relevant	0	1	1
Ratios			
Relevance 1 & 2			
Precision	100	50	
Base Recall	100	100	
Matched Recall	100	100	
Relevance 1			
Precision	100	50	
Base Recall	100	100	
Matched Recall	100	100	

RELEVANT BASE DOCUMENTS

Relevance 1 Relevance 2

429431

SEARCHER MODE

A CL

B NL

QUESTION 5-05

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	Relevance 2	Relevance 3	
1	437549	CNM		✓		T
2	[illegible]	CN	✓			T
3	430221	C			✓	
4	426505	CN	✓			T
5	422704	CN	✓			
6	419695	CN	✓			
7	323976	CN			✓	
8	313050	CN	✓			
9	430448	N			✓	
10	330477	N	✓			T
11	319981	N			✓	C
12	434360	N		✓		
13	432331	N			✓	
14	429427	N	✓			
15	427358	N		✓		
16	414787	N			✓	
17	410947	N		✓		
18	331993	N		✓		
19	330470	N			✓	
20	326487	N			✓	
21	325096	N			✓	
22	321959	N			✓	
23	318887	N			✓	
24	421685	M		✓		CN
25	431536	C		✓		N

ANALYSIS

	C'L.	N.L.
Items Retrieved	16	44
Overlap	15	
Search time	64	45
Pre-search time	5	15
Total sets	39	55
Search sets	20	33
Combine sets	19	22

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	5	7	3
Relevance 2	2	5	1
Non-relevant	2	10	0
Ratios			
Relevance 1 & 2			
Precision	78	54	
Base Recall	100	100	
Matched Recall	54	100	
Relevance 1			
Precision	55	32	
Base Recall	-	-	
Matched Recall	71	100	

RELEVANT BASE DOCUMENTS

Relevance 1

Relevance 2

421685
434549

SEARCHER MODE

A CL + NL

B NL + concept

QUESTION 5-06

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	Relevance 2	Relevance 3	
1	433661	C	✓			
2	415904	C			✓	
3	330211	C			✓	
4	421892	CM		✓		
5	415895	C		✓		
6	430029	C		✓		
7	422259	C			✓	
8	416949	C			✓	
9	411991	C			✓	
10	430569	N	✓			
11	421891	NM		✓		T
12	413599	N	✓			T
13	413595	N	✓			T
14	415554	C		✓		
15	330211	C			✓	
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						

ANALYSIS

	C'L.	N.L.
Items Retrieved	11	4
Overlap	0	
Search time	77	67
Pre-search time	45	10
Total sets	23	59
Search sets	16	42
Combine sets	7	17

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	1	3	2
Relevance 2	4	1	1
Non-relevant	6	0	0
Ratios			
Relevance 1 & 2			
Precision	45	100	
Base Recall	50	50	
Matched Recall	55	44	
Relevance 1			
Precision	11	75	
Base Recall	—	—	
Matched Recall	25	75	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
	421892
	421891

SEARCHER MODE

A NL

B CL

QUESTION 5-07

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	434652	NC	✓			T
2	423150	N		✓		C
3	430592	N	✓			C T
4	425757	N	✓			C T
5	327541	N		✓		C T
6	313174	NC	✓			T
7	433672	C	✓			
8	431695	C			✓	
9	431678	C			✓	
10	431055	C			✓	
11	426663	C			✓	
12	415863	C			✓	
13	414910	C	✓			
14	413855	C			✓	
15	411943	C		✓		
16	332672	C			✓	
17	327725	C			✓	
18	326160	C			✓	
19	324250	CM	✓			N
20	322126	C		✓		
21	316534	C			✓	
22	429586	M	✓			CN
23	412880	M	✓			CN
24	413857	M		✓		C
25	320211	M	✓			CN

ANALYSIS

	C'L.	N.L.
Items Retrieved	81	21
Overlap	19	
Search time	43	30
Pre-search time	45	5
Total sets	63	20
Search sets	40	13
Combine sets	23	7

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	5	4	4
Relevance 2	2	2	1
Non-relevant	10	0	0
Ratios			
Relevance 1 & 2			
Precision	42	100	
Base Recall	100	80	
Matched Recall	100	63	
Relevance 1			
Precision	30	67	
Base Recall	100	100	
Matched Recall	100	71	

RELEVANT BASE DOCUMENTS

Relevance 1: 324250, 429586, 412880, 320211

Relevance 2: 413857

SEARCHER MODE

A NL + CONCEPT

B CL + NL

QUESTION 5-08

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	Relevance 2	Relevance 3	
1	430549	C	☑	✓		
2	430530	CM	✓			N
3	430529	C		✓		N
4	420810	C		✓		N
5	415834	C		✓		
6	323271	C		✓		
7	312143	C		✓		N
8	432515	N		✓		
9	430556	N		✓		
10	430524	N		✓		
11	422834	N		✓		
12	420809	N		✓		C T
13	415869	N		✓		
14	411230	N		✓		
15	410501	N			✓	
16	332046	N		✓		
17	331106	N		✓		
18	330131	N		✓		T
19	329084	N		✓		T
20	326426	N		✓		
21	324219	N			✓	
22	319195	N			✓	
23	313187	N			✓	
24	312151	N		✓		
25	311142	N		✓		C T

ANALYSIS

	C'L.	N.L.
Items Retrieved	15	47
Overlap	8	
Search time	15	38
Pre-search time	41	54
Total sets	25	31
Search sets	16	23
Combine sets		

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	1	0	0
Relevance 2	6	14	4
Non-relevant	0	4	0
Ratios			
Relevance 1 & 2			
Precision	100	75	
Base Recall	100	100	
Matched Recall	43	85	
Relevance 1			
Precision	14	0	
Base Recall	100	100	
Matched Recall	100	100	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
430530	

SEARCHER MODE

A NL

B CL

QUESTION 5-09

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	421194	C		✓		N
2	315937	CN	✓			
3	434499	C			✓	
4	421339	C		✓		N
5	425554	C		✓		
6	420177	C			✓	
7	410281	C			✓	
8	325654	C		✓		N
9	425764	C			✓	
10	415804	C			✓	
11	329593	C		✓		
12	330562	C		✓		N
13	311648	C			✓	
14	432047	N			✓	
15	430552	N			✓	
16	426758	N			✓	
17	426098	NM		✓		C T
18	421194	N		✓		C
19	417920	NM	✓			C
20	414910	N			✓	
21	333148	N			✓	
22	330360	N		✓		C T
23	327609	N			✓	
24	325405	N			✓	
25	427075	M			✓	

ANALYSIS

	C'L.	N.L.
Items Retrieved	54	46
Overlap	15	
Search time	79	25
Pre-search time	5	10
Total sets	55	17
Search sets	32	14
Combine sets	23	3

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	1	2	1
Relevance 2	6	3	2
Non-relevant	6	7	0
Ratios			
Relevance 1 & 2			
Precision	54	42	
Base Recall	100	100	
Matched Recall	100	81	
Relevance 1			
Precision	8	16	
Base Recall	100	100	
Matched Recall	100	100	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
417920	426098

SEARCHER MODE

A NL + Concept

B C.L. + N.L.

QUESTION 5-12

RELEVANCE DECISIONS

	Document	Mode	Relevance 1	2	3	
1	434537	CN	✓			T
2	431524	C			✓	
3	420677	C	✓			
4	415710	C	✓			
5	410013	C			✓	
6	330050	C			✓	
7	325292	C			✓	
8	321037	C			✓	
9	312308	CN			✓	
10	434716	N	✓			C T
11	433433	N			✓	
12	428782	N			✓	C
13	422941	N			✓	
14	417026	N			✓	
15	414984	N			✓	
16	410342	N			✓	
17	329151	N			✓	
18	326027	N	✓			C T
19	314958	N	✓			
20	310310	N			✓	
21	434715	M		✓		N
22	422625	CM		✓		N
23	430415	CM	✓			N
24	431500	M	✓			N
25	433951	M	✓			N

ANALYSIS

	C'L.	N.L.
Items Retrieved	58	76
Overlap	37	
Search time	74	17
Pre-search time	10	15
Total sets	39	25
Search sets	22	22
Combine sets	17	3

PERFORMANCE

	C.L.	N.L.	Title
Totals			
Relevance 1	4	4	3
Relevance 2	1	0	0
Non-relevant	6	9	0
Ratios			
Relevance 1 & 2			
Precision	45	30	
Base Recall	40	100	
Matched Recall	87	75	
Relevance 1			
Precision	36	30	
Base Recall	33	100	
Matched Recall	85	71	

RELEVANT BASE DOCUMENTS

Relevance 1	Relevance 2
433951	422625
431500	434715
430418	

SEARCHER MODE

A NL

B CL

APPENDIX D

FAILURE ANALYSIS SHEETS

Question 1.01

A difficult question. User knew of no relevant documents, nothing found in manual search and no Relevance 1 documents. In spite of the two searches retrieving 76 and 49 references, only 6 were in common. With Controlled Language, five of the retrieval sets were N.L. and one was jointly N.L. and C.L.

Assessment	N.L. and A.C. File	
	Rel. 2. Mismatch (3)	

Question 1.02

Straightforward question, with appropriate C.L. terms. No Relevance 1 failures.

Assessment	N.L.	C.L.
		Rel. 2. Indexing (4)
	Rel. 2. Abstract (1)	

Question 1.03

Straightforward question. C.L. all Natural Language terms. Both searches would have retrieved 2 of the user-given relevant documents on titles.

Assessment	N.L. and A.C. File	C.L. and N.L.
	Rel. 2. Search (2)	Rel. 1 Search (1)
		Rel. 2 Search (2)

Question 1.04

Straightforward question, but difficult to ensure that only relevant documents were retrieved. Relatively low overlap of total retrieved, but those that were retrieved by both systems were relevant. Simpler search to structure by N.L. but took far longer.

Assessment	N.L.	C.L.
	Rel. 1. Search (2)	Rel. 2 Search (2)
	Rel. 2. Abstract (2)	
	Rel. 2. Search (2)	

Question 1.05

Diffuse question, with possibly relevant documents dealing with seemingly unrelated matters. However, user asked for a 'noisy' search to give him large output to evaluate but this was ignored by N.L. searcher, who selected a set with only two references. C.L. search a mixture of Controlled Language and Natural Language.

Assessment	N.L. and A.C. File	C.L. and N.L.
	Rel. 1 Search (1)	Ref. 2 Search (1)

Question 1.06

Difficult question to hold to requirements. No known relevant documents. N.L. searcher only retrieved one document against 90 by C.L. searcher. N.L. search failure due to 'caesium' being spelt as 'cesium' in abstract.

Assessment	N.L	C.L.
	Rel. 1. Search (1)	Rel. 2. Mismatch (2)

Question 1.07

Reasonable question but failure in searching to distinguish between 'separation' (aerodynamics) and 'separation' (mechanical). This might have been avoided by specifying NASA classification. C.L. search mainly Natural Language terms.

Assessment	N.L. and A.C. File	C.L. and N.L.
	Rel. 2. Search (2)	Rel. 2. Mismatch (1)
		Rel. 2. Search (1)

Question 1.08

Apparently reasonable question, but some documents not well indexed. For example a relevant document was indexed as 'decimeter waves' instead of 'decameter waves', and in another case was not indexed as such although the term occurred in the title. N.L. had difficulty in eliminating thermal emission.

Assessment	C.L.
	Rel. 1. Indexing (2)
	Rel. 2. Indexing (1)
	Rel. 2. Search (2)

Question 1.09

Apparently reasonable question, but searches failed to distinguish between descriptions of and scientific objectives of L.S.T., and this resulted in high retrieval of non-relevant documents. C.L. search mainly Natural Language.

Assessment	C.L. and N.L.
	Rel. 1. Search (1)
	Rel. 2. Search (1)

Question 1.10

An apparently precise question, but one which was, in fact, either very broad or was generously assessed. Of 41 documents found by either C.L. or N.L., all of which are presumed to be relevant, only 4 were found by both systems. This reinforces the implication of the low base-recall ratio that there were probably between 100 and 200 relevant documents in the data base. Most of the recall failures by N.L. were due to the requirement that the term 'modal', with no alternative, must appear, but the C.L. failures could mainly be ascribed to incomplete indexing.

Assessment	N.L.	C.L.
	Rel. 1. Abstract (1)	Rel. 1. Indexing (3)
	Search (2)	Rel. 2. Indexing (2)
	Rel. 2. Abstract (2)	Search (2)
	Search (4)	
	Data Base (2)	

Question 2-01

Question probably not suitable for NASA STAR. Difficult to find appropriate index terms, and C.L. search was almost entirely N.L. The only Relevance 1 paper retrieved (by C.L.) appears to have little relation to the question as stated.

Assessment	N.L. & A.C File	C.L. and N.L.
	Rel. 1 Mismatch (1)	Rel. 2 Searching (4)
	Rel. 2 Search (4)	

Question 2-02

Apparently reasonable and precise question, but few references retrieved, although questioner could cite a number of relevant papers.

Assesment	C.L.
	Rel. 2 Searching (1)

Question 2-03

General question with apparently large number of relevant papers. C.L. search mainly index terms, but over-specific and failed to include aspect of pollution.

Assessment	C.L. and N.L.
	Rel. 1 Indexing (1)
	Searching (3)
	Rel. 2 Searching (3)

Question 2-04

A straightforward question, but C.L. required a large number of index terms in a simple search statement.

Assessment	C.L.
	Rel. 1 Searching (1)
	Rel. 2 Searching (4)
	Indexing (3)

Question 2-05

A reasonable question, but in both C.L. and N.L. presented some difficulty in bringing in the requirement for low Reynolds number, thus causing more non-relevant items than might have been expected.

Assessment	N.L. & A.C. File	C.L. and N.L.
	Rel. 2 Searching (1)	Rel. 1 Searching (2)
		Rel. 2 Searching (1)

Question 2-06

Straightforward question, but the general term 'errors' can cover a wide range of different types or causes. The C.L. searcher felt it was 'a good search for C.L. because the thesaurus gave all the related terms', and this was borne out by the comparative result. Precision failures due to a mismatch, as questioner failed to specify that her interest was in equipment used in commercial aircraft.

Assessment	N.L.	C.L.
	Rel. 1 Abstract (1)	Rel. 1 Indexing (1)
	Rel. 1 Searching (5)	Rel. 2 Searching (2)
	Rel,2 Abstract (1)	
	Searching (2)	

Q2-07

Question 2-07

Reasonable question, needing few search terms. N.L. search failed to retrieve several relevant papers by requesting 'programmable (F) Calculators' as against calculators, C.L. search used all natural Language terms and had the simplest search of the whole test (i.e. 1+2+3+4+5)

Assessment.	N.L. & A.C. File	C.L. and N.L.
	Rel. 1 Searching (6)	Rel. 2 Mismatch (1)
	Rel. 2 Searching (2)	

Question 2-08

A reasonable question, but major relevance papers more likely to appear in journal articles.

Assessment	N.L.	C.L.
	Rel. 2 Searching (3)	Rel. 1 Indexing (1)
		Searching (1)
		Mismatch (1)
		Rel. 2 Indexing (2)
		Searching (5)
		Mismatch (1)

Question 2-09

Appears to be a reasonable question, but apparently poor results by both searches. Most non-relevant items dealt with control of sub-systems and not, as required, control of whole satellite in orbit. However, questioner had not expected that many papers would be found, and the relavance 1 document that was retrieved 'highlighted work of which we were not aware'.

Assessment	N.L. & A.C. File
	Rel. 2 Searching (1)

Question 2.10

Good question. N.L. searcher expressed surprise at number of documents with 'subsonic' or 'low speed', but this aspect was frequently missed in indexing and this accounted for most of the C.L. failures. Relatively low base recall ratios indicate there were probably many more relavant papers.

Assessment	C.L.
	Rel. 1 Indexing (4)
	Searching (2)
	Rel. 2 Indexing (4)
	Searching (4)

Question 4.01

A reasonable question, with high level of relevant documents (mainly Relevance 2) and well suited to C.L. index terms (e.g. 'Moving target indicators)

Assessment	N.L. & A.C. File	C.L. and N.L.
	Rel. 2 Searching (2)	Rel. 1 Indexing (1)
		Rel. 2 Indexing (2)
		Searching (3)

Question 4.02

An apparently reasonable question, but low precision ratio due to inability to meet the requirement regarding applications of aerosol optics.

Assessment	N.L.	C.L.
	Rel. 1 Data base(1)	Rel. 1 Searching (1)
	Rel. 2 Searching (1)	Rel. 2 Mismatch (2)

Question 4.03

A difficult question to match, as shown by the fact that there was no overlap between C.L. (30 refs.), N.L. (32 refs) amd Manual Search (4 refs). Precision failures largely due to macro-climate and not micro-climate.

Assessment.	N.L. & A.C. File	C.L. and N.L.
	Rel. 2 Abstract (2)	Rel. 1 Search (2)
	Mismatch (1)	Rel. 2 Search (3)
	Search (1)	Indexing (1)

Question 4.04

A difficult question, probably not suited to NASA Star

Assessment	N.L.	C.L.
	Rel. 1 Searching (1)	Rel. 2 Searching (2)
	Rel. 2 Searching (1)	Mismatch (1)

Question 4.05

A (technically) simple question, but not suited to NASA Star, and searches understandably failed to retrieve anything really applicable.

Assessment	C.L. and N.L.
	Rel. 2 Searching (1)

Question 4-06

A general question, with probably many more relevant papers than were retrieved. Only 7 terms used in N.L. search, giving maximum precision but low recall. C.L. failures due to the many possible terms covering 'computers' allocated to relevant documents (e.g. computer programs, computer system design, computerized design, computer design)

Assessment	N.L.	C.L.
	Rel. 1 Searching (2)	Rel. 1 Indexing (3)
	Rel 2 Searching (4)	Searching (2)
	Mismatch (4)	Rel. 2 Indexing (1)

Question 4-07

A good but awkward question. Successful N.L. search with only five terms. Unusual in that nine of eleven N.L. relevant documents could have been retrieved on titles, probably because searcher restricted two terms to titles only, and thereby failed to retrieve four relevant papers.

Assessment	N.L. & A C File	C.L. and N.L.
	Rel. 1 Searching (1)	Rel. 1 Data base (1)
	Rel, 2 Searching (2)	Searching (3)
		Rel. 2 Indexing (3)
		Searching (1)

Question 4-08

A difficult question to cover all possibilities, with many different reasons for precision failures. Easier to structure by natural language.

Assessment	N.L.	C.L.
	Rel. 2 Data base (1)	Rel. 1 Searching (1)
		Indexing (1)
		Rel. 2. Indexing (1)

Question 4-09

Straightforward question. Simple N.L. search (four terms) retrieved large number of relevant documents.

Assessment	N.L. & A.C. File	C.L. and N.L.
	Rel. 1 Searching (1)	Rel. 1 Searching (3)
		Data base (1)
		Indexing (1)
		Rel. 2. Searching (1)

Question 4-10

Straightforward question for C.L., the searcher acceptingall documents indexed by any of four terms (e.g. decompression sickness), and index terms covered concept not mentioned in abstract. N.L. search should have accepted 'decompression' with other terms.

Assessment	N.L.		C.L.	
	Rel. 1	Searching (3)	Rel.1	Indexing (1)
	Rel. 2	Searching (2)		

Question 5-01

Difficult question to match. Most of the non-relevant items did not deal with high-level diffusion. C.L. search mainly natural language. Long and complex N.L. search failed to retrieve a number of relevant documents.

Assessment.	N.L. & A.C. File		C.L. and N.L.	
	Rel. 1	Searching (2)	Rel. 1	Searching (1)
	Rel. 2	Searching (5)	Rel. 2	Searching (1)

Question 5-02

An apparently simple question, but the C.L. searcher insisted on an association of inertial navigation with missiles. This additional concept may have come up in discussion with the questioner, but the many C.L. failures were mainly due to the lack of such index terms in the relevant documents found by N.L.

Assessment	N.L.		C.L.	
	Rel. 1	Data Base	Rel. 1	Searching (5)
				Data base (1)
			Rel. 2	Searching (7)
				Indexing (1)

Question 5-03

C.L. search concentrated on means, N.L. on objects, but neither successful for this awkward question. C.L. search entirely natural language,

Assessment	N.L. & A.C. File		C.L. and N.L.	
	Rel. 1	Searching (1)	Rel. 1	Mismatch (2)
	Rel. 2	Searching (2)	Rel. 2	Mismatch (2)
		Mismatch (1)		Searching (1)

Question 5-04

A precise question, with relevant items found by display of large set of possible items. Although successful in this way, it should have been simpler to find with C.L. indexing. No comparative failures.

Question 5-05

A reasonable question. As C.L. used mainly natural language, the difference in output is due to the more specific search requirement of C.L.

Assessment	C.L. and N.L.	
	Rel. 1	Searching (2)
	Rel. 2	Searching (4)

Question 5-06

From the titles of known relevant papers, this is a question with wide range of possible search terms, but with the possibility of high non-relevant retrieval. Both searchers kept retrieval low, with no overlap.

Assessment	N.L.		C.L.	
	Rel. 1	Searching (1)	Rel. 1	Mismatch (1)
	Rel. 2	Searching (4)		Indexing (2)
			Rel. 2	Mismatch (1)

Question 5-07

Good search question, for which N.L. achieved high precision by insisting on 'adaptive'.

Assessment	N.L. &	A.C. File
	Rel. 1	Searching (1)
		Mismatch (1)
	Rel. 2	Mismatch (2)

Question 5-08

A more general question than it seemed, with apparently a large number of related papers but little specifically on the subject. Difficult to ascertain reasons for failures, being a mixture of mismatch, indexing and searching.

Assessment	N.L.		C.L.	
	Rel. 2	Mismatch (1)	Rel. 2	Mismatch (5)
		Searching (2)		Searching (7)

Question 5-09

Fair question, with comparable output by both modes, with C.L. search being mainly natural language.

Assessment	N.L. &	A.C. File
	Rel. 2	Searching (1)
		Abstract (1)

Question 5-10

A tricky question, due to probability of false coordination between vortex wake and vortex turbulence. Precision failures in N.L. due to use of turbulent wakes.

Assessment	N.L.		C.L.	
	Rel. 1	Searching (2)	Rel. 1	Indexing (1)