CHAPTER 5

Formation of Index Languages

The indexing described in the last chapter provided a number of different index languages: first, one consisting of single terms in the natural language of the documents indexed; second, this initial language made more precise by the recognition of 'concepts', reflecting a first level of interfixed relations; third, a yet more precise language recognizing a further level of interfixed relations in the form of 'themes'; fourth a language in which the relative importance of the terms was recognized (in the form of weights). Combinations of these provided still more precise languages; e.g., combination of the third and fourth.

Insofar as the indexing recognized substantive or lexical elements primarily and lacked the relational, or syntactic device of role indicators, it was something less than completely exhaustive. But apart from this, all the precision devices had been accommodated and the next step was to establish facilities for expanding the elementary classes and forming further index languages - i.e., to construct recall devices. This chapter deals with this activity in relation to
    1. Single terms
    2. Simple concepts
    3. A pre-established thesaurus

1. Single-term classes

A preliminary task was to prune the natural language indexing of certain minor inconsistencies and variants which had inevitably crept in and which were not in themselves regarded as sufficiently serious methods of defining classes to warrant separate measurement. These initial controls involved the following:
(1) Singular and plural forms were confounded;
(2) American and English and other variant spellings were confounded; e.g. gage and gauge, fiber and fibre, Von Karman and Karman.
(3) Certain qualifiers of terms (affixes, hyphenated-forms which were sometimes separated, etc.) were disregarded; e.g., built-up, pitch-up, rolled-up, etc. were treated as built, pitch, rolled; ellipse-like, jetlike, etc. were treated as ellipse, jet.
(4) Numbers as qualifiers were separated and treated as separate terms; e.g. Mach 6 became 'Mach' and '6', N.P.L. 18 x 4 (a wind tunnel) became 'N.P.L.' and '18 x 4'.

Table 5.1 gives the basic data regarding the number of single terms and their frequency of use after the above preliminary controls had been imposed. The full set of indexing terms is given in Appendix 5.1

Salient points are: for a collection of 1,400 documents the total vocabulary was 3,094 terms, with reductions to 2,668 and 1,816 for the less exhaustive vocabularies, (the reduction being based on the weights assigned to each term). The average number of terms used to index a document was 31.3, reduced to 25.2 and 12.9 respectively for the less exhaustive vocabularies. (A discussion of the problem of reduced vocabularies appears below).

As to the use of different terms, whilst the average number of times a term was used was 14.2 this is not a very significant figure in view of the wide scatter. Of the 3,094 terms, 1,169 were used only once; one term (Flow) was used 942 times, another (Pressure) 720. The distribution curve for word-use is shown in Table 5.2 where it is compared with three other indexes, with larger vocabularies. It can be seen that the distribution behaves as expected in view of the fact that it reflects a smaller vocabulary than the other three. In fact, the frequency of use proved to be remarkably consistent with the well-known Zipf distribution of words according to their frequency of use in natural language texts. It will be seen that some 10% of the terms (the most

Collection size   1400 documents
Total postings of terms   43,857
Average postings per document   31.3
Total unique terms   3094

Variations in exhaustivity

	Total terms in vocabulary	Average Postings per document
Maximum exhaustivity (all weights)	3094	31.3
Medium exhaustivity (Weights 7/10)	2668	25.2
Minimum exhaustivity (Weights 9/10)	1816	12.9

Use of terms

Average usage per term	14.2
Terms used once only	1169
Terms used more than once	1925

The first ten terms, ranked by usage:

Flow (942)
Pressure (720)
Boundary (512)
Layer (512)
Distribution (442)
Theory (400)
Velocity (360)
Supersonic (352)
Mach (344)
Equation (312)

Variations in vocabulary size (according to different index languages)

Language 1 (Natural language, single terms only)	3094
Language 2 (Lang. 1 with synonyms confounded)	2988
Language 3 (Lang. 1 with word forms confounded)	2541
Language 4 (Lang. 1 with synonyms and word forms confounded)	2444
Language 7 (Lang. 1 with minimum hierarchical reduction)	1217
Language 8 (Lang. 1 with medium hierarchical reduction)	796
Language 9 (Lang. 1 with maximum hierarchical reduction)	306

(383 Proper names are not included in the counts for languages 7, 8 & 9)
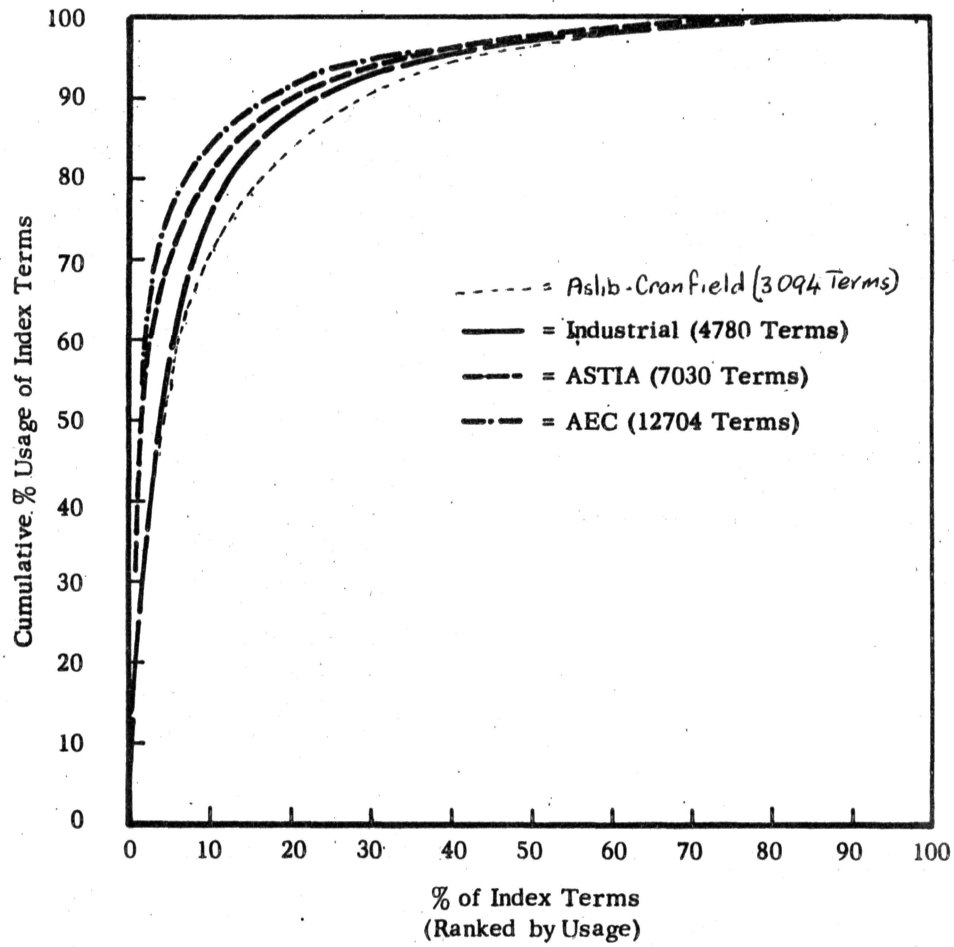
FIGURE 5.1 NATURAL LANGUAGE SINGLE TERM DATA

FIGURE 5.2 PATTERNS OF TERM USAGE

(original
(Original figure taken from 'Centralisation
and Documentation", Arthur D. Little Inc)

used ones) accounted for 68% of the indexing postings, and 30% accounted for 92% of the postings, after which the curve flattens out.

Reduced vocabularies

Some explanation of the problem of vocabulary reductions referred to above seems desirable. Generally speaking, all recall devices imply a smaller vocabulary (with bigger classes), and precision devices imply a larger vocabulary (with smaller classes). A class is enlarged by confounding two or more classes which previously had a separate existence; contraction is the reverse process. By 'vocabulary', we mean the total number of discrete indexing elements, lexical and syntactic (i.e., substantives and relational terms) provided in an index language. It may seem surprising that links are included in a statement of vocabulary size, since they are not discrete devices in the sense that they are countable in the way lexical terms and roles are, but vary with the number of documents indexed. However, by the fundamental criterion of whether they define particular classes which would not be distinguished without them, they must be regarded as part of vocabulary size.

It should be noted that vocabulary size, under normal indexing conditions, is not necessarily a determinant of the specificity possible in an index language. This is because increased specificity is always obtainable by coordination; e.g., if the vocabulary contains the terms Flow and Supersonic, class Supersonic flow is specifiable by coordinating these two terms. Theoretically it is possible to specify almost anything in this way; e.g., Air x Cushion x Vehicle is a simple conjunction of the separate terms normally used to name this thing; but even where a name in no way defines the nature of the thing it represents, it may be specified uniquely by contrived analytical 'definition' e.g., in the W.R.U. Semantic Code, Tempering is represented by Process x Metal x Heat x (number) where the number is an arbitrary code symbol distinguishing this particular heat process on metal from any other. Perhaps the extreme example of the use of reduced vocabularies, with precise description resting on the various conjunctions of a few fundamental terms was the Malvern experiment (Ref. 25).

In the case of single-term classes without coordination, however, a reduced vocabulary can be an absolute bar on the specificity possible. If no coordination is used, a single-term vocabulary of 1,500 specifies only half the classes specified by a 3,000 term vocabulary. So far as testing devices is concerned, there are two different ways of effecting the expansion of classes. One is by an absolute reduction of vocabulary whereby the reduction is obligatory for all searches; the other is by selective search programmes, whereby the effective reduction is permissive and may or may not be utilized in a particular search. In the first case the reduction is measurable (i.e., in terms of the number of discrete classes distinguishable) and in the other it is not.

Obligatory reduction of vocabulary

Here, there is an absolute 'block reduction' (a block of classes being condensed into one) in the number of classes recognized, and the indexer and searcher has no option but to accept the confounding of more specific classes which is implied. This was the case with reduction by synonym-control and by confounding of word forms. It was also the case with the single-term hierarchies, although reduction by hierarchy may be achieved permissively and was in fact done this way in the testing of 'concept' hierarchies. This point is explained later on.

The use of quasi-synonyms* to enlarge classes is a permissive device; since the final, expanded classes do not totally exclude the continued separate use of the terms confounded (as is the case with real synonyms) no figure showing the exact degree of vocabulary reduction is possible. For example, the expanded class for Bow is Bow + Bowing + Ahead + Front + Forward + Forebody; the expanded class for Ahead is Ahead + Forward + Upstream. Clearly, the expansion of Bow does not result in the obliteration of the separate class 'Ahead', which not only continues to exist but is in turn expandable by the addition of other quasi-synonyms. In an index language which confounds true synonyms only, the reduction is once and for all and the terms no longer have a separate identity.

At this stage of our thinking about the function of vocabulary size as the main determinant of recall and precision, it seemed desirable to have as exact a measure of this parameter as possible. So the first testing of hierarchy took the form of a fixed (and therefore accurately measureable) reduction in vocabulary size.

Hierarchical reduction

The two methods of measuring hierarchy as a recall device (i.e., by obligatory, block reductions in vocabulary size and by selective searching through different hierarchical paths) are demonstrated below. The first example is one of a 'concept' hierarchy - i.e., one not restricted to single terms and one place per term. This is in order to show more clearly the two methods, and also to emphasize the distortion which results from restriction to a 'one-place' hierarchy of single terms. The latter results in the exclusion of some terms which are located in more general categories and the result is seen in the second part of the example. A hierarchical notation is attached to this example in order to make the permissive search clearer (in the schedules actually used, notation was purely ordinal).

'Concept' hierarchy demonstrating hierarchical reduction

a	Experimental wind tunnel methods for investigating flow
ab	Visualization methods
aba	Using smoke, vapours, etc. (3)
abaa	Vapour screen (1)
abab	Fog (1)
abac	Wood smoke (1)
abad	Oil smoke (1)
abb	Using coatings, flows, etc. (3)
abba	Oil flow (1)
abbb	Oil film (1)
abbc	China clay (1)
abbd	Phosphorescent lacquer (1)
abbe	Ink flow (1)
abc	Using spectrum (3)
abca	X-ray spectrography (1)
abd	Using Stroboscope (3)

---

*'Quasi-synonyms' are terms which can be used synonymously in certain contexts, but which are not true synonyms. (see page 68)

[Experimental wind tunnel methods]
　　[Visualisation methods]

abe	Using shadowgraph (3)
abf	Using photography. Photorecording (3)
abfa	Drum camera (1)
abfb	Photomultiplier (1)
abfc	Television (1)
abfd	Motion picture (2)
abfda	High speed (1)
abfe	Schlieren (2)
abfea	Spark (1)
abfeb	Photomultiplier (1)
abff	Interferometry (2)
abffa	Fringe shift (1)
abffb	Interferential strioscopy (1)
	etc.

The full hierarchy provides twenty-nine classes.

The **first** reduction, replacing the terms marked (1) by a see reference to their immediate containing head (e.g. Drum camera see Photography), leaves eleven classes, namely

a	Experimental wind tunnel methods
ab	Visualisation methods
aba	Using smoke, vapours etc.
abb	Using coatings, flows etc.
abc	Using spectrum
abd	Using stroboscope
abe	Using shadowgraph
abf	Using photography
abfd	Motion picture
abfe	Schlieren
abff	Interferometry

The second reduction, similarly replacing terms marked (2), which now include those originally marked (1), by see references to their containing heads, leaves eight classes, namely

a	Experimental wind tunnel methods
ab	Visualisation methods
aba	Using smoke, vapours etc.
abb	Using coatings, flows etc.
abc	Using spectrum
abd	Using stroboscope
abe	Using shadowgraph
abf	Using photography

The third reduction similarly replacing terms marked (3), leaves two classes, namely

a	Experimental wind tunnel methods
ab	Visualisation methods

In this way figures are obtainable to show the exact effect of moving, say, from

the quite specific class Interferential strioscopy to Interferometry in general, then to Photographic methods in general and then to Visualization methods in general, and so on.

It will be noted that by the second reduction there is no separate class left for Schlieren photography, whilst a distinct class is retained for Stroboscope. Yet in this subject field Schlieren photography is a decidedly more important class than Stroboscope   This suggests that reduction by purely hierarchical criteria may be unsatisfactory. When we reflect that the choice of terms within categories and the choice of the categories themselves is ultimately a matter of literary warrant, it is reasonable to assume that reduction of classes hierarchically should not be a rigid process, but should take note of the weight of literature in the different classes, so that Schlieren photography, for example, might be retained although all other sub-classes at that level were removed and incorporated in the containing class. For the single term hierarchies, this line of reasoning led to the abandonment of 'pure' hierarchical reduction and the incorporation of judgements as to the relative importance of particular classes, and the noting of word frequencies in determining which classes should be retained intact at a particular level of reduction.

The above example stresses the primary function of hierarchy as a recall device, whereby the index vocabulary is systematically reduced and the scope of each remaining class is consequently widened (hence the greater recall). In practice, however, by varying search programmes, hierarchical linkage allows movement in both directions - to greater precision by refining class definition or to greater recall by coarsening class definition. If specific indexing is assumed (i.e., each document, or document-theme, is assigned to its most specific class) a search may be made in a number of different directions through the hierarchy. For example, a searcher commencing at Photorecording abf may find the amount of material there unexpectedly excessive and so decide to search a narrower class. This may be done, of course, independently of hierarchy, by adding a qualifier or two (e.g., moving from Photorecording to Photorecording in high speed wind tunnel). But it may also be done by moving down the hierarchy (e.g., to Schlieren tests).

Such a decision implies that the links established by the hierarchy are permissive, not obligatory, and that the searcher selects from a comprehensive hierarchy just those routes he considers likely to be fruitful. It is not certain whether figures produced for a number of such searches would be useful in the sense of allowing firm generalizations to be drawn since much will depend on the subject field, and on the choice of pathways followed in searches (4) and (5) as discussed on the next page. Assume that, as shown below a, aa, etc. are terms in hierarchical relation:

```
a
  aa
    aaa
    aab
    etc.
  ab
    aba
    abb
    abc
    etc.
  ac
  ad
  ae
  etc.
```

Then, if a request is made for ab, there are the following alternative basic programmes which can be used.

(1) Term and species - i.e. ab + aba, abb ... abz.

(2) Superordinate - i.e. ab + a (but excluding other subclasses of a; this is how 'generic' search is popularly interpreted in general library practice where a might represent a general treatise on the subject of which ab is a subclass).

(3) Generic - i.e. a + aa + aaa, aab ... aaz + ab + aba, abb ... abz + ac + aca, acb ... acz + az + aza, azb ... azz (this is how 'generic' search is normally interpreted in machine systems and is analogous to search (1) where the content of class ab is taken to include the individually specified subclasses aba, abb, abc ...).

(4) Coordinate - i.e. a selection of the more likely classes coordinate with ab, e.g. ab + aa + ad; e.g., in a category of three-dimensional shapes a search for Spheroid can be extended generically by searching under Body of Revolution, or by examining all the different kinds of Body of Revolution (Sphere, Hemisphere, Ogive, Cone, etc.). But some of the latter will be more closely connected to Spheroid than others (e.g. Sphere) and an intermediate search, stopping well short of examining every species, can be made. It is true, of course, that a 'closer connection' between several subclasses implies the possibility of an intermediate step of division being inserted. But we have to stop somewhere.

(5) Subordinate - i.e. a selection of the more likely subclasses of ab, e.g. abb + abn.

It will be noted that programmes (1), (2) and (3) are obligatory; no freedom of choice is given to the searcher, but (4) and (5) are permissive, the decision as to the formation of the classes being at the discretion of each searcher.

To return to the matter of variations between single-term and concept hierarchies, the shrinking of a concept hierarchy by restricting it to a one-place hierarchy of single terms is seen by the following, which is the schedule given on pages 62-3 and reduced in this way.

V/2	Experiment + Experimental
V9a/10	Visual + Visualization
V11	Spectrography
V12	Stroboscopic
V13	Shadowgraph
V14,16	Photography + Photorecording
V17	Schlieren
V18	Spark
V19	Interferometry
V21	Interferogram
V23	(Fringe) Shift
V24	Strioscopy
V25	Interferential
V25a	Clay
V25b	China

Of the 29 classes in the concept hierarchies only 13 appear in the one-place hierarchy plus two (Interferogram and Interferential) which appear for the reasons

explained on p. 70. The others are now distributed under the more general categories as explained: i.e., terms like Smoke, Vapour, Screen, Fog, etc. appear in other contexts as well and are therefore placed in more general categories. Under these conditions, as soon as reduction of the original full vocabulary begins, it becomes very difficult to maintain the sensible boundaries of a class like Visualization tests. For if a question on this were now programmed to included such terms as Fog, etc., it is very likely that these in turn have been swallowed up in the reduction of the general categories and that their inclusion in the search programme can only be had at the cost of bringing in a number of other terms, such as Cloud, Snow, etc. which are quite irrelevant to the context of Visualization tests.

Another drawback, related to the foregoing, is the loss of connection suffered by terms treated in isolation and not in coordination. For example, a search in response to a question on 'Flow in channels' would fail to draw in documents indexed by 'Couette flow' or 'Poiseuille flow'. Although there is a clear connection between these at the 'concept level' of types of flow, at the level of single terms there is no connection between Channel (treated as a Structure affording a passage) and the personal names Couette and Poiseuille. This situation reflects a practical difficulty in post-coordinate systems which rely on single terms - that of indicating connections (in a thesaurus, say) when these connections are dependent on particular conjunctions; e.g. this would imply a reference of the kind:
Channel: when coordinated with Flow
    see also: Couette Flow
                    Poiseuille Flow.

It is important, therefore, to remember that the performance results of the single-term hierarchies reflect the use of one particular application of hierarchy as a recall device - i.e., its expansion of classes by fixed reduction of vocabulary size. Also, that this was a procedure determined largely by considerations of measurement rather than regard for the normal use of hierarchy as a recall device in practical indexing. There seems little doubt now that it is a mistake to regard hierarchy as an obligatory recall device. Its essential function is to act as a permissive device, allowing flexible choice of class adjustment according to the demands of the question context in a way which is not feasible within the artificial conditions of single-term hierarchies. From this viewpoint, the performance figures for the concept hierarchies described in the next section are a better guide to the value of generic hierarchy as an indexing device.

Languages based on single-terms and embodying recall devices

Before describing these in detail it may be noted that a certain artificiality inevitably accompanies the application of recall devices to single terms in isolation, simply because, in many cases, words make little sense when stripped of accompanying qualifiers, etc. For example, the problem of synonymity in index languages frequently demands recognition of phrases, as when 'Ground effect machine' is equated with 'Air cushion vehicle' although at the single term level there is no synonymity between the constituent terms; and a term like 'effect' on its own is practically valueless as a retrieval handle (which is what any class, in indexing, aims to be).

Traditional, pre coordinate indexing has always begun with some degree of coordination. Even in analytico-synthetic classifications, where 'elementary constituent terms' are separated out as far as possible, there is no rigid adherence to the single term as the basis of the language; for example, 'Ground effect machine' would be comfortably accommodated in a Vehicles facet. But coordination of terms is an extremely

potent precision device and whilst the measurement of its impact, alone and in conjunc-
tion with other devices (including all the recall devices) was of course essential, it
could not be included as a variable when measuring the impact of the other devices
on single terms. Completely free manipulation of classes is only feasible if we begin
with single terms; this is a basic assumption of post-coordinate systems. It was
clearly desirable to obtain performance figures for the impact of single devices on
single classes before attempting to measure the joint impact of several devices - and
even a slight degree of pre-coordination would have compromised such figures.

Confounding of synonyms

This is perhaps the most obvious of all indexing devices and the one least likely
to be neglected even in the crudest of indexes. Much of this work was straightforward:
e.g., recognition of synonymity between such terms as Acoustics and Sound, Amount
and Quantity, Calculation and Computation, Axisymmetric and Axisymmetrical, Vertex
and Apex, Viscid and Viscous. However, exact synonymity is relatively rare (there
might even be argument about some of the examples above). The commoner situation
is a partial synonymity, where terms are interchangeable only in particular contexts.
The evident richness of the English language, even in the literature of high-speed aero-
dynamics, led to quite different terms being used on different occasions (but often in
the same document) to represent the same thing; e.g., the notion of Proximity might
be conveyed by that term or by Near, Nearest, Nearly, Close, Closely, Off, Adjacent,
Contact, etc. Two terms which might be used synonymously on most occasions would
occasionally diverge seriously; e.g., Interplanetary flight is equated with Interplanetary
voyage; Hypersonic flight with Hypersonic flow, Free flight with Free falling. But
Voyage, Flow and Falling cannot be regarded as synonyms.

The establishment of a synonym-list suffered one unfortunate drawback in that
it preceded the construction of classification schedules. Ideally, a synonym-list
in any given area should be extracted from a detailed classification; only by a system-
atic organization of all used terms according to their meanings can the ramifications
of complete and partial synonymity be exposed. For administrative reasons, however,
it was desirable to proceed with the measurement of relatively straightforward devices
like synonyms, word-forms, weights, etc., whilst the preparations for the more dif-
ficult devices like hierarchical linkage went on.

The truth of the assertions just made was borne out when the classified hierarchies
were completed, in that a number of further synonyms, unrecognized in the synonym
programme, were disclosed. However, these cases were relatively few and we are
satisfied that the synonym-list on which the tests were made was reasonable on the
whole.

One difficult decision necessary in establishing the synonym list was whether
we should recognize variant word forms as synonyms. Whilst the usual view of syno-
nymity excludes variant word forms as being examples of a grammatical rather than
a semantic relationship, the practice of many subject heading lists, thesauri, etc.
which fail to recognize variant word forms at all is an implicit acceptance of the view
that such variants are virtually synonymous. Certainly, in the process of indexing
by natural language terms extracted from the documents, the fact that one word form
rather than another was selected was often almost fortuitous and this is shown, with
examples, in the section on hierarchical linkage. However, this argument was not
regarded as acceptable; a thesaurus, etc. may fail to recognize variant generic levels
as well as variant word forms, and so implicitly confound a genus and its species.

Should a synonym list accept this also as a type of synonymity? In view of the fact that separate measurement was being made of these other devices it was decided to interpret synonymity as strictly as possible, but the joint use of synonyms and word-forms (for example) was also measured.

Another weakness, showing itself as a result of the single term basis, has already been referred to - the inability to cater for synonyms which appear only at the 'concept' level of phrases; e.g., Flexural centre and Shear centre, Surface friction drag and Skin friction, Uniform surface temperature and Constant wall temperature.

These considerations led inevitably to the recognition of 'quasi-synonyms' as a variant of 'pure' synonymity.

Confounding of quasi-synonyms

In this device, those terms are confounded which on some occasions, but not all, are used synonymously; e.g. Subsonic and Subcritical; Compressor, Impeller and Pump; Blunt, Blunted, Bluff and Rounded; Medium, Environment and Surrounding; Region, Atmosphere and Material. A certain overlap appears here with the device of confounding word-forms; on many occasions, different word-forms would be used in a report indiscriminately to convey the same notion. The same overlap would appear, of course, with 'true' synonyms if 'conveying the same meaning' were the sole criterion. But in a single-term index language, the extra element of context is lacking; although the phrases 'Seasonal density variation' and 'Variation of density with season' were virtually synonymous in the reports indexed, if the single terms 'Seasonal' and 'Season' are taken alone they cannot be regarded as synonymous.

With quasi-synonyms this restrictive rule did not apply, since acknowledgement of differences conveyed by variations in context is the basis of the device. So variant word-forms were accepted, where applicable, as one type of quasi-synonym, e.g. Flexural and Flexible, Flow and Flowing.

The establishment of synonyms and quasi-synonyms was done as the indexing progressed, with the aid of glossaries, classification schedules, etc. We have already noted that theoretically, the only sure way of tracing synonyms is by a close classifiction of the field, utilizing the defining functions of classification to expose synonymity between terms used. However, the compilation of classification schedules was a much longer task and, for clerical reasons, the list of synonyms was compiled first, in the manner indicated above. Again, for clerical reasons, the synonyms were worked out fully only in the case of those which were demanded by the search programmes - i.e., the starting terms from the questions.

Confounding of word-forms

There is little to be said of this device, which was the simplest of all to establish. The expanded classes consisted of a comprehensive aggregation of all the various forms a given word-root could take, whether with prefixes, suffixes, participles or gerunds, etc. Examples are: Angle, Angled, Angular and Angularity; Asymptote, Asymptotic and Asymptotically; Axial, Axially and Axis; Blunt, Blunted, Blunting and Bluntness; Bound, Boundary, Bounded and Bounding. Their relations to synonyms and quasi-synonyms have been mentioned and their place in one-place single term hierarchies will be considered in the next section.

The terms used in searching, together with their synonyms, word endings and quasi-synonyms, are given in Appendix 5.2.

One-place, single-term hierarchies

By far the most difficult device to establish was that involving hierarchical linkage. Two major (and connected) problems arose. Firstly, the arbitrary and somewhat artificial restriction implicit in the need to place each term in one hierarchy only. This arose inevitably from dealing with single terms and meant that the assistance normally given to definition by context was absent. Secondly, the problem of interpreting the prolixities and ambiguities of the natural language index vocabulary in terms of this particular type of controlled vocabulary.

Problems of hierarchy

The term hierarchy as normally used in indexing can mean one of three different things:

1) A generic hierarchy; i.e., a system of subordinating some terms to others whereby only terms which reflect the relationship of being kinds of a thing are subordinated to that thing. Other relations are excluded. But the basis for the formation of the species may or may not be a 'fundamental' characteristic.

2) A strict genus/species hierarchy, differing from (1) in that it is confined to the use of 'fundamental' characteristics; e.g., Methane could not be subordinated to Fuel (as it is in the test schedules) since a fundamental definition of Methane does not require characterization by this attribute. A parallel has been drawn by Gardin (Ref.26) with the distinction between paradigmatic and syntagmatic relations, the former reflecting permanent or necessary relations and the latter temporary or contingent ones. A modern faceted classification uses both types of hierarchy in that the same term might appear in two or more different facets according to its status (as a product, an agent, etc.) and not be confined to the facet where it 'fundamentally' belongs.

3) A hierarchy which includes generic and non-generic elements; i.e., one which subordinates some terms to others regardless of the relation involved, so long as the subordinated term can be seen to belong to some category or facet of the 'containing' class, e.g., the subordination under a term of its properties, parts, processes, etc., as well as its kinds. This situation is typical of nearly all existing library classifications.

Reasons why (3) should be treated as a separate device ('non-generic hierarchical linkage') have already been given and are not considered here. In choosing between (1) and (2) for single-term hierarchies, logic seemed to suggest that (2) be chosen; for if each term may go in one hierarchy only it is arguable that that one place should at least reflect the most essential characteristics of the thing represented. On the other hand, the practical purposes of hierarchy in indexing would sometimes be ill-served by such an arrangement. This purpose is to provide for each term a set of class-mates standing in the same relation (of Thing/Kind) to the containing class and thus facilitate the expansion, or contraction, of any given class by the inclusion or exclusion, of some or all of these helpfully related neighbouring terms; and 'helpfully' here depends on the subject context.

If the terms are relegated to a 'fundamental' or 'common' category, these helpful relations tend to become tenuous; e.g., if the term Upper is located in a highly

generalized category of Spatial Phenomena, its class mates will eventually include such terms as Underwater or Buried. In a question on the Upper atmosphere, if expansion of the first term brings in such classes it is clearly unhelpful. Many other examples could be given; e.g., Boosted and Reinforced could reasonably be assigned to a basic category of activities affecting the dimensions of a thing. But in the collection indexed these terms appeared in quite different contexts - Boosted under Rocketry and Reinforced under Structures.

The procedure finally adopted was a compromise. Where, in the test collection, an index term had appeared only in one particular context it was placed in a hierarchy reflecting that context, without regard to whether it was a necessary or contingent relation. For example, Gun would be regarded as a method of propulsion in any fundamental hierarchy; but in the test collection it appears only as a designation of a kind of aeronautical testing device (a special kind of wind tunnel) and so it was located with the latter. An extreme example would be Gamma; as a single term, this could hardly appear in any 'fundamental' category other than Letters; in the test collection it appeared only as the designation of a kind of steel and was located as such.

If, however, a term appeared in several different contexts suffering a significant qualification of meaning, it was placed in a 'fundamental' category; e.g. Integral appeared in its mathematical and structural sense and was therefore placed in a Common properties category. Similarly, the term Working appeared in two main guises: to designate a particular section of a wind tunnel and to designate a fluid (e.g., a test gas). The sense of the term Working is significantly different in the two contexts and it was therefore relegated to a common properties category.

Problems of terminology

Closely related to the above problem was that of interpreting the intended meaning (from the point of view of the test collection) of the terms used in the natural language indexing. The organization of terms into hierarchies constitutes a form of controlled vocabulary, of course; in this case, it was a control being exercised retrospectively, after the indexing stage. The object was to place each term in the hierarchy to which it would have been assigned had the indexing been done using the controlled vocabulary. So where the same essential notion was conveyed in various grammatical styles, these variants would have been ignored and one form done service for all; that is to say, the particular grammatical form of a term might have to be disregarded since its semantic content in the index was the only point of interest now. For example, a writer might refer indifferently to 'reduction of x by compression', or 'reduction of x by compressing' or 'reduction of x compressively' without wishing to convey a significantly different idea. Again, any one of the phrases 'plate with curvature', 'curved plate', 'curve of the plate', and 'curving the plate' might be used in a report without any intention of conveying different nuances of meaning (i.e., without meaning to refer to the structure or the property or the operation in particular). Other examples were Test and Testing; Calculation, Calculating and Calculated; Asymptote, Asymptotic and Asymptotically. All these variations in expression were ignored and the different forms juxtaposed in the hierarchies.

Where different word forms reflected significantly different emphases in meaning they were assigned to their formal categories, So Buckle and Buckling appeared as processes and Buckled as a property; Cantilever was used to characterize a kind of beam, but Cantilevered designated a type of structure. Scooped appeared as a property and Scooping as a process.

In the same way, there were numerous examples of terms which appeared to represent operations or processes (if one regarded only the single terms in isolation) but which represented an integral part of the specification of a particular kind of thing; e.g. Settling chamber, Driving gas, Non-lifting wing, Geared elevator. Wherever such a term had appeared only in that particular context and its function as a class determinant had been to characterize the entity and not the operation, **property**, etc., as such, it was subordinated in the hierarchy to the entity which it specified.

The exact status of these variants on insertion into the hierarchies created a slight, theoretical problem. The confounding of synonyms in an earlier programme had already established what terms were exactly synonymous and it would have been inconsistent now to add these variants as synonyms (the weakness of a synonym programme derived before the establishment of a classification has already been noted). So they were simply clustered together as though coordinate in relation to each other. Had the measurements of single-term hierarchical linkage taken the same form as in the later 'concept hierarchies', whereby various hierarchical trails were followed in order to distinguish sharply between different relations (subordinate, superordinate, coordinate, etc.): this might have produced a very slight distortion of the performance figures. However, the measurement of single-term hierarchies only took the form of block-reductions in vocabulary size (in the manner discussed earlier in this chaper), so no harm was done.

It must be admitted that a few errors crept in, when unjustified violence was done to a category by the subordination of one of its members to another category. For example, in the overwhelming majority of cases, the term Revolution occurred in indexing as part of 'Body of Revolution'; so, according to the reasoning above, it was located in the category of Shape, since its function was to designate a particular kind of shape. However, its synonym, Rotation, occurred once or twice in its fundamental guise of a process; it is therefore misplaced under Shape. It is not thought that these occasional lapses were serious. We have already seen that in making single term hierarchies, if a term is relegated to a fundamental category this results in classes sometimes being drawn in which are unhelpfully associated; this is also what happens in the case of a lapse like the above.

Construction of single term hierarchies

Having settled on the various solutions to the problems described above, the formidable task of organizing the 3094 terms of the natural language proceeded. The basic operation was one of facet analysis (a facet being a hierarchy). A useful framework for the initial sorting was the Facet Classification compiled for the first Aslib-Cranfield Project by J. Farradane and B. C. Vickery, although high speed aerodynamics (the subject of this test collection) tended to concentrate itself in only a few of the areas covered by the scheme, and was in far greater detail than had been handled before. Particularly large categories were those relating to Bodies, to Shapes, and various Spatial and general relations, to Fluid dynamics proper, with particular clusters of detail under such topics as Compressors, Upper atmosphere studies, and Astronautics. The speed with which the last subject has developed in recent years was reflected in the fact that whereas the Facet Classification barely mentioned it, in this test collection it was a major theme.

Because no attempt was made to establish 'fundamental' categories as such, the common categories which were formed tended to be residual ones in that they contained only those terms which had not found a place in a more limited context.

A1	Particles
A3	Electron + Beta
A4	Proton
A5	Atom
A6	Isotope
A7	Ion
A8	Molecule
A8a	Mol

	(Structure)
A8b	Atomic
A8c	Molecular
A8d	Bimolecular
A9	Homonuclear
A9a	Nuclear
A10	Monatomic
A11	Diatomic
A12	Polyatomic
A13	Polymer
A14	Polycrystalline

A15	Matter + Material
	(By use)
A16	Pigment
A17	Lacquer
A18	Phosphorescent
A19	Ink
A19a	Injectant
A20	Fuel
A21	Methane
A22	Ethylene
A23	Hydrocarbon
A24	Methanol
A26	Propellant
A27	Explosive
A28	TNT
A30	Coolant
A31	Lubricant
A32	Refractory
A32a	Oxidizer
	(By origin)
A34	Electrodeposit
A35	Electroformed

	(By constitution)
A38	Metal
A39	Alloy
A39a	Bimetallic

FIGURE 5.3 SAMPLE SHEET FROM SCHEDULES OF
SINGLE TERMS

Nevertheless, some of them were still exceedingly large and detailed - e.g., those reflecting spatial and shape characteristics. For these, and for the common categories of Properties, Processes, Operations, etc. the Thesaurus and Code Dictionary FROLIC produced at the David Taylor Model Basin (Ref.28) proved very useful.

Interpretation of the various word forms, etc. referred to above was assisted by the file, compiled during indexing, of synonyms, definitions, decisions, etc., by the word frequency list, and by reference to the indexing sheets of individual documents where necessary.

Sample excerpts from the single-term hierarchies are given in Fig. 5.3. (The complete schedules appear as Appendix 5.3.) It must be emphasised that only those terms appear which were used in the indexing of the test collection. Whilst this resulted in very detailed schedules in some areas, these still cannot be regarded as exhaustive of the terms in the particular area. Sometimes, if they did not happen to occur in the test collection, quite important terms will be missing.

2. Simple concepts

In the previous section we described the establishment of index languages based entirely on single words, and indicated the limitations on the performance of synonyms and hierarchies imposed by this restriction. These limitations were accepted in order to allow the examination of the performance of the different devices applied to single terms, in the absence of any element of precoordination. The next step was to accept a degree of precoordination from the outset.

Examples have already been given of the sort of simple linking necessary if the meanings of some expressions in the natural language are not to be quite lost; e.g., 'Ground effect machine' must be retained as a single concept if loss of meaning is not to be suffered. The original indexing had, of course, included a statement of the 'concepts' in each document - it was in fact the first step taken in the actual procedure of indexing a document. These concepts were now taken as the basis for the production of new synonym and hierarchy languages.

'Concept' languages

In order to reduce the task of preparing these to reasonable proportions it was decided to take a substantial subset of the full collection of 1400 documents and to make a detailed classification schedule for all the terms appearing in it. The subset consisted of some 200 documents, containing all the documents relevant to some 40 questions. In order to make the new collection reasonably homogeneous, only aerodynamics documents were included.

The performance of the index languages in this same subset was subsequently measured separately for a controlled language (based on a thesaurus) and for the 'options' investigated by G. Salton and his colleagues at the Harvard Computation Laboratory (the SMART system). Figures for the single term languages for the subset had already been obtained - they had simply to be extracted from the figures for the full collection.

No reindexing was attempted, of course, since this would have invalidated comparisons with the single-term tests. One adjustment was made, however; the

original concepts, based closely on the natural languages of the documents indexed, reflected a degree of precoordination which was excessive for our purpose, so over-elaborate phrases were now broken up into smaller units, e.g., Biconvex circular arc cross section was broken up into Biconvex cross section and Circular arc cross section; Dissociated frozen hypersonic laminar boundary layer became Dissociated boundary layer, Frozen boundary layer, Hypersonic boundary layer and Laminar boundary layer. If search were subsequently necessary for the original concept, it would still be possible by postcoordination. Meanwhile, this splitting up allowed maximum freedom in distinguishing facets and subfacets (arrays). In other words, the rigidity attending the excessive precoordination typical of the older classification systems (resulting in the obscuring of the multiple relations between facets and sub-facets) was avoided.

Formation of concept hierarchies

This task proceeded in the normal way, by the now well-established process of facet analysis. However, some of the problems which occur when making a special classification were absent or greatly reduced; at the same time, the unusual basis of the schedules (the 'natural language' concepts, already embodying a certain degree of precoordination) raised some new problems of presentation. These points are discussed later.

The procedure was as follows: the concepts (mostly short phrases like Tumbling entry, Centre of rotation, Crossed flexure pivot, but with some single words, e.g., Strips, Trajectory, Pivot, Inclination) were first grouped into the following major subject areas:

> Aircraft types and parts
> Bodies (Aerodynamic)
> Non-aerodynamic structures
> Flight: flying operations
> Fluids, gases, atmosphere
> Fluid flow: Kinds, Elements (vortices, jets, etc.)
> Aerodynamic forces and loads, processes and properties
> Aeroelasticity, flutter
> Aerodynamic reference parameters (angle of attack, planform, etc.)
> Mechanics, dynamics
> Heat
> Research: Experiment, Theory
> General properties and processes.

No particular significance attached to this order; for convenience of reference it approximated to the order of terms in the original Cranfield Facet Classification. Generally speaking it reflected the citation order used in locating concepts; a concept containing notions from more than one area was located under the one appearing first in the above sequence; e.g., Wing-body interference went under Wing-body, not Interference; Spherical segment nose went under Nose (Aircraft parts) not Spherical segment (Bodies); Leading edge stall went under Leading edge. But where a clear relation, explicit or implicit, existed between two elements of a concept, and reflected a clear precoordinate indexing principle (e.g., subordination of agents to the operations or processes they serve) this was observed, even if it ran counter to the broad rule above; e.g., a Shielding mechanism is a structure (non-aerodynamic)

but it was subordinated to Ablation cooling since its functions in the literature indexed was that of an agent of the cooling process.

The fact that Heatshield was subordinated to Ablation cooling devices did not mean, of course, that it was unavailable as a member of the class Structure if this latter subordination had also been required. It was placed under Ablation solely on the score that in the test collection, or the subset, this was its most probably useful hierarchy. The concept schedules were essentially 'one-place' schedules in linear sequence, in the sense that no attempt was made to repeat one concept in several different positions should it happen to belong usefully to several different hierarchies. This last event was provided for by the rotated A/Z index, and by references within the schedules described later. It must be emphasised that the function of these schedules was simply to show as clearly and as comprehensively as possible the hierarchical relations (generic and non-generic) between the terms (concepts) so that searches could be programmed from them. The major relations were most economically displayed by physical juxtaposition. Other hierarchical relations were established via the A/Z index and by internal references (linking, for example, Heat transfer, subordinated to Thermodynamic processes, to Transport properties in general).

The index in the physical sense (the matrix of index descriptions and document numbers) consisted of the separately entered concepts designed to be searched post-coordinately (the strip, or the scan-column, method used is described in Chapter 6). So a compound like Fully developed laminar channel flow could be sought equally in any of the various arrays concerned, or combinations from them - Fully developed flow, or Laminar flow, or Laminar channel flow, etc.

It follows from the above that the problem of citation order was very much reduced, compared with a real life schedule, since it was confined entirely to the choice between two (and sometimes, but rarely, three) elements; e.g. Jet noise, Interference rocket, Laminar boundary layer heating, Surface stress, Slot blowing. But whilst these particular examples offered a serious choice between two or three equally important elements, the great majority did not even demand this; they consisted of combinations such as High pressure ratio compressor, or Hinged flap, where the major element was obvious and the other elements trivial; no hierarchy of Ratios, or things High, or things Hinged was necessary. Since the concepts represented the limits of precoordination, the problem of providing for the accurate prediction of the exact location of all potential synthesized combinations (a major function of citation order) did not arise. The problem of 'distributed relatives' was solved by postcoordination of the concepts.

It also follows from the above that problems of notation were virtually non-existent. A purely ordinal notation to identify quickly the location of particular simple concepts was the only requirement. Problems of hospitality and expressiveness did not arise; no additions to the schedules were contemplated and no aids to display were necessary in schedules which were relatively homogeneous and designed for internal test-programming entirely.

Within each major area the various facets and arrays (subfacets) were now distinguished. At this point the problem of displaying generic and non-generic hierarchies arose; it was met as it usually is in conventional library classification, by subordinating to a thing its various categories - its kinds, parts, properties, processes, etc. Below is a brief extract from the schedules which are given in full in Appendix 5.4, followed by an explanation of some of its features:

E64   Compressor
E65   Centrifugal + Radial flow compressor.
                          + Radial flow turbomachine
E66   Axial flow c. + A.c. + A.f. turbomachine
E67   Drum construction
E68   Disk construction

E69   Axial flow compressor blade
E70   Naca 65 (12) 10 Blower blade

E71   Jumo 004

E72   Single stage compressor
E73   Multi stage compressor

                (etc. - i.e., other kinds of compressor)
                [Parts]
E89   [Stage] Q.
E90   Stage characteristic
E91   S. efficiency + S. performance
E92   Cascade losses
E93   S. matching
                    (etc. - i.e., other Stage characteristics)
                [Blade]
F19   Rotor blade
F20   Stator blade
                    (etc. - i.e. other kinds of Blade)
                  (Blade characteristics)
F35   Blade shape
F37   B. curvature
                    (etc.)
                [Flow phenomena]
F91   Irrotational flow
                [Rotational flow
F92   Inlet whirl
F93   Prewhirl
                (etc.)

The first subclasses under Compressor are <u>Kinds</u> of compressor: Centrifugal and Axial flow (based on direction of flow): Single stage and Multi-stage (based on stage numbers) and so on.   The synonyms which appear at the concept level automatically sort themselves out (e.g., the three variants at E65).   Any categories (generic and non-generic) which refer to a given subclass follow that subclass immediately.   So under Axial compressors is found <u>Kinds</u> (Drum construction, Disk construction, Jumo 004) and <u>Parts</u> (Blade) - and a particular kind of axial compressor blade follows that. (N.b. - a clerical error has resulted in the Kind of a.f.c. 'Jumo 004' following the Part, 'Blade' instead of preceding it; such errors did not affect the programming of searches).

      The <u>Kinds</u> facet is followed by the <u>Parts</u> facet (Stage, Blade, etc.).   The bracketed term Stage followed by 'Q' indicates a term which appeared in one of the questions but not in the indexing of the subset documents; it has been inserted for programming purposes.   The array of 'Stage characteristic' demonstrates a recurrent problem in the subject field analysed, that of maintaining 'generic' relations in a situation where strict definition of terms would result in an uncomfortably large number of tiny subfacets

consisting often of a sole member; e.g., Stage efficiency is evidently a property and Stage performance a process - but they are treated as virtually synonymous; Cascade losses constitute a factor in efficiency or performance, but hardly a 'kind of efficiency'. Stage matching is another concept which lies on the borderline between processes and properties. It is possible to say, however, that all these rather subtly related notions are 'Stage characteristics' and in this way the facet structure is maintained without undue complexity. Other examples of a certain amount of violence being done to the strict nature of generic relations may be found, as at M4/33 (see Appendix 5.3) where complexly related terms are grouped as Atmosphere properties and characteristics, or at P7/26 Processes and properties of Vortices. Similar situations inevitably arose in the single-term hierarchies in areas like Mechanics and Dynamics where the conditions of what Ranganathan has called 'canonical' classification tend to hold.

Another minor liberty, not demonstrated in the example above, was taken in the treatment of qualifying terms like Theory, Approximation, Experimental data, when these were found precoordinated as in Hypersonic flow approximation. It could be argued that these do not narrow the extension of the term they qualify and should therefore be disregarded - i.e., treated as synonymous with the term alone. Theoretically, this is why they are usually placed (in the guise of 'Form divisions') at the very beginning of the subdivisions of a term in conventional classification. In the search programmes, however, they were included in the 'Terms and species' sub-programme. This was later seen to be a mistake, but it is not thought that this was serious in view of the very few terms involved.

Although the differently related facets follow and interrupt each other without clear signs of demarcation in the schedules, the different relations were strictly observed, of course, when the search programmes were compiled; i.e., when expanding a class by generic hierarchy, only those terms standing in a true generic relation to that class were counted; e.g., Compressor + Centrifugal c. + Axial flow c. + Drum construction + Disk construction + Jumo 004 + Single stage c. + ... would be given as the full generic expansion of a particular kind of compressor (see the Generic - Broad search below). Any terms not standing in a true generic relations (e.g., Axial flow c. blade, Stage characteristics, Irrotational flow, etc.) would be ignored.

Multiple hierarchical relations

The major weakness of the linear display of classes just described, in which a particular class (concept) is located in one place only (albeit a carefully chosen one) is that it fails to show the further generic relations a class may have. For example, Jet interference is subordinated to a category Jet characteristics, in which its class mates are Jet exit, Jet location, Jet energy, Jet structure, Jet emission, etc. It could equally well be subordinated to a category of Causes of interference with class mates like Wake interference, Forebody interference, Support system interference, Wave reflection interference, Wing-Body interference, etc. But in the schedule described these last terms are 'distributed relatives'.

There are two traditional methods of handling this problem in a real-life classified index: by multiple-entry as with UDC, where the number of entries for a given compound-class-description are multiplied, and filed according to a different classification, so that Jet interference appears in a class Jet (divided into Jet characteristics) and also in a class Interference (divided into Causes of interference). Or, by leaving these other connections to be indicated by an A/Z relative index, in which all the

different contexts in which a term appears are gathered together as qualifiers of that term.

Although in a real-life situation the first method provides easier access to these further relations, this advantage was not significant in the test environment. The test collection subset was relatively small and a fully rotated A/Z index was easily producable by clerical labour. In any case, such an index was necessary for other reasons, too, as will be described. Moreover, although the working out of full alternative hierarchies would have involved a considerable effort, there was no guarantee that more than a small fraction of them would ever be used, since only those hierarchies relevant to the terms occurring in questions would be required.

The assumption above is that such an A/Z index will automatically disclose the existence of other possible hierarchies. Indeed, it is difficult to see how such additional hierarchies could be economically developed unless we are guided by the literary warrant afforded by the actual occurrence of the terms concerned in conjunction with these other contexts, in which case the A/Z index automatically picks them up. Nevertheless, further connections were indicated by references in the schedules wherever it seemed desirable, particularly where it seemed that the A/Z rotation of terms might still fail to show the connection; e.g. Small disturbance theory, subordinated to Disturbance, contains a reference to Boundary layer theory to which it is also relevant. Streamlines, subordinated to Flow elements, has a reference to Relative stream surfaces (in Compressor flow phenomena) to which it is also generic. Or, within a given class, references were added to link concepts occurring in different arrays; e.g., Performance discontinuities in the Performance facet of Compressors contains a reference to Stall and Surge in the Flow facet of the same class.

It may be noted that in a real-life classified index, the A/Z index usually shows even those connections just exemplified, since its entries contain more qualifying material (providing further information regarding the context) than the test index, where the concepts gave the sole element of precoordination. For example, a document dealing with Small disturbance theory in the context of Boundary layer theory would produce rotated A/Z index entries:

> Small disturbance theory: Boundary layer theory
> Boundary layer theory: Small disturbance theory

and these establish the connection which in the test collection had to be established by references.

The significance of multiple hierarchical linkage as an element in the recall performance of generic hierarchical linkage generally is probably not very great. Most questions impose a particular context of their own and the likelihood of relevant material being found in radically different contexts of the particular terms of the question is probably small. For example, a question on the kink in the surge-line of a multi-stage axial compressor imposes a context on the notion of 'surge'; clearly, documents indexed under Surge as a general concept should be examined, but it is unlikely that extended examination of the classes flanking Surge in the general hierarchy of Aerodynamic processes would be very fruitful.

The A/Z Index (see Appendix 5.5)

In order to provide for multiple generic hierarchical linkage as discussed above, and for other reasons, a rotated A/Z index of the concepts was now produced; e.g.,

	Class No.
Afterbody, Conical Base	
,Cylinder	
,Cylindrical	
Drag	D6
Surface	D5
,Truncated	
Vehicle, Conical	
(etc.)	
Base Afterbody, Conical	
Bleed	T16
,Flat	
Forward attitude	K61
(etc.)	
Cone	H55
Cone, Blunt Nose	
,Blunted	
,Circular	
Cylinder	H76
Cylinder Bodies	H76
(etc.)	
Conical Afterbody Vehicle	A57
Base Afterbody	C99
Camber	W33
(etc.)	

It can be seen that each concept appeared as many times as it had distinct words. So the first concept above appeared in three different contexts - that of Afterbody, of Base and of Conical. The class number appeared after the direct form of the concept.

The index served the obvious purpose of a key to location besides its other major purpose - that of indicating all the different contexts in which a given term had appeared in the schedules. One aspect of this second function, the capacity to reveal other generic hierarchical relationships, was discussed above. But this was only one kind of context revealed. In the example above, Afterbody surface, Base bleed, Base forward attitude, etc. reflect non-generic relations. The index therefore acted as a valuable supplement to the schedules proper in displaying these relations. The major display of these was, of course, by the subordination of a thing's categories to that thing. But these would not necessarily exhaust the non-generic relations, and the A/Z index not merely supplied further relations, but could lead the question programmer back into the systematic order to explore further categories, if necessary. For example, examining the entries adjacent to Heat transfer leads to Heat sustaining leading edge (subordinate to Leading edge), to Heat transfer at the wall (subordinated to Surfaces and Walls, where related concepts such as Constant wall temperature and Wall temperature gradient are found) and to Heated air (subordinated to Air, where related concepts such as High temperature air and Dissociated air are found). Many of these other concepts do not contain the term 'Heat' or its variants and might not have been picked up had purely alphabetical considerations governed the search.

A third major function served by the rotated A/Z index was to provide a recall device based on the 'accidental' alphabetical juxtaposition of concepts enjoying a limited

degree of precoordination. It has already been shown how the cluster of concepts around a given term (which might also be a root term for a number of word forms) such as Heat or Interference reflects a variety of relations; e.g., Interference, Blockage; Interference, Forebody; Interference, Jet; these all reflect kinds of interference according to source. Interference filters reflects Interference as an experimental agent (in temperature measurement): Interference load reflects Interference as a source of another phenomenon. When these assorted relations are added to a certain degree of word-form confounding, (e.g. expanding an initial enquiry for Dissociation by the addition of classes like Dissociated stream or Dissociating fraction) the result is an eclectic recall device which utilizes elements of hierarchy, non-generic hierarchy, confounding of word-forms, and linking (an element of precoordination is essential to the programme). Such a mixture cannot, however, rank as a 'device' in the way this notion was understood in Chapter 4. It is further considered in the next section.

Formation of Classes by Search Programmes

A significant feature of hierarchical linkage as an indexing device is the rich variety of relations it displays, enabling a number of different paths to be pursued in adjusting the size and content of the class or classes with which a search begins. Some of these paths were briefly mentioned in the last section, using the example of Visualization tests.

In exploiting these relations two different policies can be followed; either classes are expanded by bringing in all the terms related in a particular way - e.g., all the terms subordinate to the original one, as when all the different kinds of compressors are added to a search for Compressors. Or, classes are expanded eclectically, choosing just those members of a given relationship which seem most likely to be relevant in the context of the whole question. The latter policy is the one normally followed in the conventional classified index.

The former policy has the merit of simplicity in programming (once the schedules are established) and this is clearly pertinent in the case of machine searching and is, in fact, generally implied by the term 'generic search'. Equally obvious is the fact that it will tend to result in a lower precision ratio than a selective search, but possibly also a higher recall ratio.

In the testing of the concept hierarchies it was decided to attempt both approaches and the following different searches were programmed, each one producing a differently defined class.

(1) The simple natural language concept alone

(2) Confounding of synonyms. It has already been pointed out that a classification should automatically throw up synonyms as a result of its analysis; also, that a number of synonyms only become apparent at the level of concepts. Both these factors operated to produce a programme for synonyms quite different from that using single terms alone. Examples are: Temperature distribution + Temperature profiles + Temperature history; Angle of incidence + Angle of attack + Arbitrary angle of attack + Incidence; Initial expansion region + Prandtl-Meyer region.

(3/8) From this point onwards, the classes formed by (2) were regarded as the basic classes to be expanded. This expansion was achieved by adding further classes to (2) on the basis of the following programmes:

(3) <u>Term and species</u>: If the basic class were Non circular cylinder and its synonyms (H75), this would be expanded by the addition of Cone cylinder + C.c. bodies + Elliptic c. + Elliptic c. of eccentricity $\frac{1}{2}\sqrt{3}$. + Hemispherical c. + C. with h. nose + Ogive c. model + Flat faced c. + C. without corners.

(4) <u>Term and species (selection)</u>: a choice was made from (3) based on the context of the question asked. For example, in a question on the kinetic theory of gases, when programming the term Gases, only those kinds of gases which reflected in some way the problem of the question were selected - such as Ideal gas, Real gas, High temperature gas, Dissociating gas, Equilibrium gas.

(5) <u>Superordinate</u> - i.e., adding to the basic class its immediate containing genus and as many more genera beyond that as appeared sensible; the number of steps included would rarely exceed three. To Non-circular cylinder (H75) would be added, for example, Cylinder + Body of revolution + 3-dimensional body. It should be noted that only the superordinate term was taken - not its species as well; the search is the equivalent of the traditional library search under 'more general' heads.

(6) <u>Generic (narrow)</u>. - i.e. adding to the basic class its immediate containing class (genus) and all the other species in the same array (subfacet) as the basic class; e.g., to Non-circular cylinder would be added Cylinder and the rest of the array based on circularity of shape, but excluding those kinds of cylinder (Inconel cylinder, Flat faced cylinder, Long cylinder, etc.) reflecting other principles of division (Material, Edge properties, Length, etc.). Similarly, if the basic class were Supersonic flow, this programme would add to it all other kinds of flow designated by speed, but excluding kinds of flow based on other principles, such as viscosity, compressibility, degree of turbulence, etc.

(7) <u>Coordinate (selection)</u>: a choice was made from (6) of the most likely terms, but excluding the superordinate term. Since by definition the classes of an array are mutually exclusive this was never a very promising search and in fact was not often productive of any terms. But in those border line situations referred to above, where the concept of generic hierarchy can only be realized practically by accepting a less-than-precise category such as 'characteristics' or 'phenomena', the likelihood was greater; e.g., in a question on Air drag the coordinate class Atmospheric rotation was accepted. Another example is that of opposites, or near-opposites, like Laminar flow and Turbulent flow, where a document frequently refers to the one even when its primary subject is the other.

(8) <u>Generic (broad)</u>: this added to (6) as many more superordinate terms as seemed reasonable, together with all their species - i.e., not just those restricted to the immediate array (subfacet) in which the basic term appeared. For example, if the latter were Supersonic flow, this search would now bring in documents indexed by any kind of flow - Laminar and Turbulent, Conical and Parabolic, Equilibrium and Non-equilibrium, etc. This somewhat undiscriminating acceptance of the complete contents of a hierarchy is the equivalent of the 'generic search' as usually understood in machine searching.

(9) <u>Systematic Collateral (selection)</u>: this was a selection from (8) analogous to the selection from (6) which produced coordinate classes (7) - again excluding the superordinate terms themselves. This search was more productive than (7) since there is often a close connection between concepts from different arrays of the same genus. This fact underlies the correlation of properties and the principle of definition by

aggregation of attributes, where a term is defined by a number of attributes, each of which reflects a different principle of division of the genus which lies at the heart of the definition, e.g., Poiseuille flow may be defined as Compressible, viscous, laminar flow between closely parallel planes - and each attribute reflects a different characteristic of division for the genus Flow. So where the basic class was Boundary layer flow, for example, related classes brought in by this programme would include Shear flow, Separated flow, Viscous flow, etc.

(10) A/Z collateral: the rotated A/Z index of concepts has already been described. This search was made first by examining the index to find the basic class (question concept) and any other concepts containing it (i.e., consisting of the basic class with further qualifications). Those which seemed likely to be relevant were now added to the basic class. For example, to the basic class Axial compressor was added Axial flow compressor blade since this included the basic class and seemed relevant. Or, to Heat transfer would be added such concepts as Convective heat transfer rate, Surface subjected to heat transfer, Laminar heat transfer distribution, etc. It may be noted that most of these further classes represent non-generic hierarchical relations of the basic class. Also, that most of the question concepts already consisted of two or more words and that in many cases there would not be any more concepts containing the one sought; e.g., this was the case with Multistage compressor, Non-circular cylinder, Dissociated stream.

For those concepts containing more than one word a 'second-level' search was also made, in which each significant word (and any of its adjacent variant word forms) was examined separately and further classes selected from the total body of concepts containing it. For example, to Axial compressor would be added Axial inlet impeller; to Surge line would be added Stall limit line and Surge. It should be stressed that these selections were made in the context of a given complete question and might vary somewhat for the same concept if the context differed. For example, in a question on the Surge line of an axial compressor, the 'second-level' for Axial compressor would reject Compressor surge (although it would be relevant to the question as a whole) because this approach was already covered by the programme for Surge line. Again, it may be noted that many of these further classes represented non-generic hierarchical relations; in addition, the combined first and second level searches generally included those terms selected from generic hierarchies in searches (4), (7) and (9) which also included the actual terms used in the basic concept.

(11) Residual hierarchical linkage. The A/Z collateral searches, although providing a large number of non-generic and generic hierarchical linkages, were restricted to those which included one or more of the terms actually used in the question. This still left a number of possibly helpful classes excluded. They could be divided into two groups: firstly all those from the non-generic hierarchies which appeared in the schedules where the question concept (the basic class) was located, but which failed to include the actual question term or terms (in which case the A/Z collateral would have disclosed them.) It was a simple matter to establish these, by scanning the various facets subordinated to a given concept, or adjacent to it.

Secondly, all related classes not already disclosed by the hierarchical relations of the ten searches described. A number of these were already provided for in the schedules, by references; e.g., Surface combustion (D66) see also Ablation; Vaneless diffuser (in compressors) (F84) see also Ducts; Compressor surge see also Rotating stall; Mass flow fluctuation (U44) see also Sound waves.

It has already been noted that the formidable task of adding to the 'one-place' schedules all other possibly useful hierarchies was not attempted. This was partly because much of the effort would have been wasted (if no questions were asked involving these alternative hierarchies), and partly because the A/Z index was likely to disclose the most important ones. It was also thought that the detailed analysis of reasons for failure (an integral part of the test programme) would disclose any examples of failure due to the absence of such alternative hierarchies.

It should be remembered that the hierarchies actually established were those reflecting the most likely approaches to the material and that for many of the concepts alternative approaches (manifested in different citation orders) were quite obviously unnecessary; to take some at random for example, Mixture of cold gases could not conceivably enter into a search for kinds of mixtures, or kinds of cold things; the same applies to a number of other concepts involving the term Mixture. Similarly, in the case of a number of concepts involving the word 'modes', or 'models', it was unnecessary to contemplate the possibility of having hierarchies based on these (although hierarchies of particular kinds of model, e.g., wind tunnel models, were of potential value, of course).

The references already provided in the schedules and by a file of 'notes and decisions' assembled during the indexing were now supplemented by those in various thesauri and subject heading lists in the field of aeronautics and astronautics, since these are in principle the product of similar observation of connections between classes. Examples of such references are those from Heat transfer to Transport coefficients, Large Peclet number and Prandtl number; from Dissociated stream to Ionized boundary layer; from Kinetic theory to Diffusion and to Transport properties. It has already been noted that all such connections could, if necessary, be incorporated in a hierarchy of the kind being tested, although no distinction was drawn between generic and non-generic relations when utilizing these references.

The combination of search programmes (10) and (11) represents, by and large, the performance of non-generic hierarchical relations largely, combined with a smaller element consisting of those supplementary generic relations not shown directly in the 'one-place' schedules. Although it has already been argued that both these hierarchical relations are generally quite secondary to the main display of generic relations, it must be regarded as a weakness of this joint presentation that separate programmes were not made for the two distinct situations.

3. Control by pre-established thesaurus

A major objective in producing the concept hierarchies described in the last section was to afford a degree of precoordination sufficient to remove the artificialities accompanying the use of single words only in the 'one-place' index language and to provide where suitable, that minimum of syntactical linkage necessary to the clear conveyance of unambiguous meaning in the index descriptions. It was thought that the resulting index language approximated more closely to the usual environment of index devices than did the first language.

By this time, the search methods developed in the course of testing the first languages were producing the first detailed performance figures for the various devices and languages concerned. Although the operation of the large number of variables produced an extremely complicated picture in that many ways of aggregating these variables and their different values presented themselves, the general picture seemed to suggest clearly enough that the performances were not very encouraging.

High recall was obtained only at a low level of precision, and as soon as the latter was improved, a precipitate drop in recall ensued.

A number of contributory causes of this were suspected. The ambiguities and inconsistencies of the language of aerodynamics suggested one. The match between the terms of the questions and the relevant documents which was, in some cases at least, very poor was another. The possibility of defective indexing was not thought to be very serious in the sense that exhaustive selection of keywords and phrases and the organization of these into concept and themes appeared to be reasonable. But a failure to recognize fully the connectivity between the terms of the languages so far established undoubtedly caused some of the failures.

Another possible factor was the unusual route by which the initial concept indexing had been translated into the different languages. In a real-life situation, this translation is done concurrently with the indexing itself, which is channelled into the controlled language as the first stage. The central elements in the test languages had so far been applied almost entirely retrospectively. Although there appeared to be no reason why this should have affected index performance, it seemed that validation of it as a method (by comparing it with a normally produced index) would be useful.

One way in which improvements in performance were thought to be possible was by putting more sophistication into the search programmes (by distinguishing between terms of different potency, between different combinations of these, and so on.) It was thought that maximum discrimination and control in searching implied the need for maximum discrimination and control in the indexing if optimum performances were to result. Again, although it was probable that the controls effected retrospectively were as valid as those imposed concurrently (as in indexing by a recognized, pre-established, control language) the slight element of doubt suggested that it would be wise to demonstrate this.

These considerations led to a decision to set up a conventional index with a different set of connectives based on a predetermined list of terms and to compare its operation with that of the natural language with retrospective controls already tested. For this, the Engineers' Joint Council Thesaurus of engineering terms, (Ref.28) was chosen as providing an up-to-date control language in the field of physical science and engineering, which contained clearly defined connectives grouped in a manner allowing convenient comparison with a number of the hierarchical searches described in the last section. A second subset of 350 documents was selected; this included the 200 documents from the first subset that was used in testing the concept hierarchies, thus allowing direct comparison with all previous programmes.

As in the case of the simple concept languages, no reindexing was contemplated, only another translation of the indexing done originally, since reindexing would have introduced an immeasurable variable; but the production of the new indexing language simulated the normal indexing situation. In this, each document is subjected first to 'concept-analysis' when it is decided what the document is about, what its significant terms are and how these are related in concepts and themes. This is followed by the translation of this information into a particular index language, with pre-established controls as to the level of specificity to be allowed and the recognition of synonyms and of other connectives between terms and between concepts.

Production of controlled index language using E.J.C.

The main problem raised by the use of E.J.C. was due to the fact that a

relatively general thesaurus was being applied to a special field. Although some loss of specificity (compared with the natural language) was regarded as inevitable, a considerable extension of the vocabulary was necessary if the specificity were not to suffer seriously. This extension raised problems of maintaining consistency with the principles with which the existing vocabulary and its syndetic structure of connectives had been developed. To assist this, the <u>Rules for preparing and updating Engineering Thesauri</u> (4th draft) November 1964 were observed as far as possible.

Selection of terms

Generally speaking, the aim was to incorporate the extra detail as unobtrusively as possible, without disturbing the distinctive character of the E.J.C. index language. The various E.J.C. methods for keeping down the size of the vocabulary were observed where feasible:

(i) Outright rejection of highly specific terms (Rule T-1) when the sense of the term could be approximated with reasonable adequacy by a broader term. E.J.C. omitted a number of prominent aeronautical and aerospace terms which did not appear to meet this criterion (e.g., Sonic boom, Tail, Stall, Bodies, Buffeting, Chord) and these were simply added. It also omitted a very large number of more precise terms and phrases occurring in the natural language indexing but which qualified for consideration under this rule. Particularly affected were those terms reflecting spatial, dimensional and temporal characteristics many of which were in adjectival form (which E.J.C. avoids); e.g., Normal, Perpendicular, Vertical, Horizontal, Behind, Outside, Below, Nearly, Large, High, Circular, Rectangular, Octagonal, Radial, Circumferential, Zero, Rate, Without, Free.

In some of these cases, where the notion was obviously dispensable because of its poorness as a retrievable handle; the term was omitted. Examples of this were Behind, Complete, Continuous, Degree, Direct, Coefficients, Effects, Horizontal, Vertical, Near, Nearly, Normal, Outer, Outside (although some of these appeared in phrases, such as Continuous loading). Outright omission was used cautiously since it diminishes the exhaustivity of the indexing. It may be noted that the main reason for holding exhaustivity constant is its effect on recall. However, the absence of a term which is completely 'non-potent' as a retrieval handle will not affect recall except in one circumstance - the use of single term searching. Theoretically, if a question includes the term Degree or Normal and this single term is searched it might retrieve a relevant document which would otherwise not be retrieved. This possibility is removed if the term is totally obliterated from the index vocabulary. However, this situation was regarded as sufficiently remote from reality to allow it to be ignored.

Strictly speaking, the only condition under which exhaustivity is affected by index language (as distinct from the personal decision of the indexer to include or not to include a notion) is when the language completely fails to provide an appropriate term even at the highest level of generality. This sometimes occurred with E.J.C. and the solution was simply to use the name of the category to which a term belonged; e.g., the term Shape was used for a whole cluster of natural language terms - Biconvex, Concave, Circular, Configuration, Diamond, Elliptical, Octagonal, Rectangular, etc. Or, the category term Position (location) was used for terms like Beneath, Outboard, Between. In this way, although specificity suffered, there was no lessening in exhaustivity.

(ii) Confounding of opposites: this was used occasionally, as in Continuum flow, Use Free molecule flow.

(iii) Avoidance of precoordination: consistency here was not assisted by the E.J.C. rules, one of which (T-1) warns against being too specific and another (T-4) warns against not being specific enough (in the matter of bound terms). One arbitrary limitation on the degree of precoordination (to 34 characters) is evidently imposed by the three-column format used in printing the Thesaurus. But only in a few cases did new combinations exceed two words (e.g., Mass transfer cooling, Blunt leading edge, Wing-Body-Tail configurations).

This policy included the representation of some concepts by an instructed co-ordination of single terms; e.g., Aerodynamic noise Use Aerodynamics x Noise (sound), Dynamic systems Use Dynamic characteristics x Systems, Sounding probes Use Sounding rockets x Space probes, Radiating body Use Radiation x Aerodynamic configuration, Reflected wave peak overpressure Use Shock wave x Reflection x Pressure. This device is not very clearly described in E.J.C. (using the ♯ and & devices) and some of the examples of precoordination make the policy no clearer; e.g., under the term Pressure is given a large number of precoordinated phrases (Pressure distribution, Pressure measurement, Pressure gradient, etc.). When a 'new' term Pressure plotting occurred, it was not clear whether to precoordinate or keep separate or confound as a near-synonym of Pressure measurement. Again, a 'new' term Circular wind tunnel might lead to acceptance by analogy with Circular saws, etc. But should Rectangular wind tunnel and Octagonal wind tunnel be similarly distinguished? Sometimes, this sort of economy in precoordination avoiding highly specific new terms, led to strange equivalents such as Root section. Use Foundations x Profile.

The record of these rejected terms and phrases, together with the ones to be used in their place, grew to large dimensions and constructed a massive 'lead-in' vocabulary from the terms and expressions of the natural language to those of the controlled E.J.C. languages. Over 1,500 entries were made for the subset, which totalled 350 documents. It should be noted, however, that a number of these rejects were simply word-form variants, e.g., Oscillatory Use Oscillations; Oscillating Use Oscillations; Oscillatory motion Use Oscillations; Elastic Use Elasticity; Edged Use Edges.

Selection of references

(1) UF (Use for) These have already been considered above as forming a lead-in vocabulary.

(2) BT and NT (Broader terms and Narrower terms) The definition of these two reciprocal relations is reasonably clear in E.J.C.

The BT reflects a true generic (Thing/Kind) relation excluding not only the obviously non-generic ones, like operations, or Properties, but also, explicitly, the Whole/Part relation, which is often loosely associated with the generic. The BT also excludes "generic families constructed on the basis of usage", so Platinum, a member of the class Metal, is not regarded as a member of the class Catalysts since it is only sometimes used as a catalyst. This seems to suggest an even stricter interpretation by the E.J.C. of the notion of 'class' - i.e., one which excludes from membership all but 'true' species in the sense that they possess permanent and fundamental characteristics, uniquely defining them.

However, this is not borne out by an examination of the Thesaurus, which suffers some inconsistency on this point. For example, Wind tunnel nozzles gives, quite correctly, Nozzles as a BT. But the previous term Wind tunnel models fails to give Models as a BT and under Models gives Wind tunnel models as a Related term (RT). i.e., a non-generic relation. Again, the term Materials has eighteen RTs; some of them are kinds of material based on Structure (Composite, Granular, etc.) some based on Properties or Behaviour (Radioactive, Magnetic, etc.) some based on Use (Structural, Molding, etc.). It would appear that the first characteristic, at least, designates true species. Again, under Plastics are listed numerous resins as NTs. The term Resins itself is treated as a synonym of Polymers, however; but no reference of any kind connects Plastic to Polymers or vice-versa, although there is limited duplication between the NTs for each of these terms. Similarly, there is no connection established between Pumps and Compressors (the latter being treated as a synonym for Air compressor) - although Pumps has numerous NTs in the form of pumps of particular application. e.g. Fuel pump.

(3) <u>RT (Related terms)</u> These are designed to show non-generic relations (as defined above) and the rules state that it is undesirable to make RTs to 'more specific' terms. However, there are numerous examples in E.J.C. of RTs which do not observe this. For example, Hydraulic equipment has numerous references to particular types of Hydraulic equipment (Hydraulic brakes, Hydraulic presses, etc.) Apart from the fact that there were cases of true species (e.g. of Hydraulic equipment) being included in the RT framework, problems arose regarding the references from terms like Quartz (as a heat shield); if we assumed that Heat shield is in the relation of RT to Quartz, should we, under Heat shield, have added Quartz as an RT? By analogy with Insulation (say), which gives as RT the material Magnesium Oxide, Quartz should have been added. But if it represents (as it does) Quartz as designating a kind of heat shield according to material, it is a 'more specific' term and such references are not encouraged.

Generally speaking, E.J.C. observes the old rule of Subject Heading lists which avoids references to adjacent headings on the score that their juxtaposition makes further reminders to the searchers (the question programmers) unnecessary. For example, there is no reference between Shock waves and Shock tubes. So this should lead to the avoidance of a reference from Ionization to Ionosphere (which was, nevertheless, made). Also under Molding materials there are references to four RTs which are adjacent entries beginning with 'Molding ...' .

Apart from these efforts to observe consistency and method in making references, the usual variety of relations appeared to be permissible and consequently RTs were added for new terms and for existing E.J.C. terms when these had inadequate connectives. Examples of the latter situation were fairly common since Aerodynamics is not a particularly favoured subject in E.J.C. For example, there are no connections between Vibrations and Elasticity, or between Supersonic flow and Shock waves; in the case of Poiseuille flow, (which may be defined as viscous, laminar flow in pipes or between closely parallel planes), the term is rejected and referred to Laminar flow but without any references linking it to the notion of pipe flow, which in this context is just as important as laminar flow.

<u>References from post coordinated terms.</u>

E.J.C. provides a number of instances in which a term is distributed between two or more wider terms; e.g., Pressure gas welding⊬ had BTs which include

'Gas welding &' and 'Pressure welding &' (the ampersands mean that the two broader terms Gas welding and Pressure welding can be jointly substituted for the single and more precise term Pressure gas welding). This procedure led to a number of similar new references; e.g., Prandtl-Meyer flow <u>Use</u> Supersonic flow x Expansion.

However, this device leads to a difficulty, inherent in post coordinate indexing, when these particular coordinations generate new reference structures of their own which are not apparent at the level of single isolated terms. E.J.C. offers no guidance on this point. For example, assume that for Conical flow the instruction is to use Cones x Flow. In this case, some sort of reference seems desirable, either from the rejected phrase Conical flow or from its constituent terms in the form:
    Cones: when coordinated with Flow, RT Shock waves.
In this particular case an intermediate connective was established by using a heading Mach cones. Examples of where the reference structure could be fairly elaborate are Flow x Parameters and Vapour x Screens x Procedures. Each of these subjects has its own set of related terms, generated entirely by the conjunction of their constituent terms, e.g. Vapour screen method, which now has its own related terms, such as Carbon tetrachloride vapour, Humidity control, Temperature control, Operational fog density.

The implied need for such connective references if the syndetic structure were to be developed raised problems of complexity in the scan-column search techniques (which were designed to be purely clerical in operation). In view of this, together with the fact that E.J.C. quite ignored such post coordinate reference needs, it was decided to follow the E.J.C. policy and rely on future analysis of searches to show where this weakness contributed to failures in performance.

E.J.C. roles

The E.J.C. system of roles which appear, without explanation, in Table 3 of Ref.28 was used in the indexing. The reasons why roles were not tested in earlier languages have already been discussed, but the availability of a ready-made set of roles seemed a useful opportunity to investigate whether some of the assumptions made there (those relating to the applicability of roles to aerodynamic literature) were justified. Unfortunately, at this stage of the project, the time factor was beginning to limit the amount of new testing which could be undertaken and it was decided that this validation was not possible.

However the decision was not made until the tentative examination of the feasibility of using the roles had been undertaken by adding them to the indexing descriptions of a small sample of reports. These, in fact provided examples of most of the objections and difficulties we had already met in the earlier attempts to use roles. An example of the difficulties inherent in using the roles may be seen if we consider a particular document and some of the questions to which it had been judged relevant. In a document (1014) on the application of piston theory and the study of aeroelastic problems, one of the themes indexed was calculation of panel flutter in supersonic flow by piston theory. Some of the problems immediately raised by the addition of roles, taking particular terms, were as follows:-

Panel is clearly a <u>patient</u> (role 9), but could conceivably be regarded also as a <u>support</u> or <u>host in a process</u> (role 5). Vibration (the E.J.C. term for flutter) is undoubtedly an <u>undesirable component</u> (role 3) in the aerodynamic behaviour of the panel.

But this is a permanent feature of Flutter in the collection context and does not alter from one document description to another. But roles are expressly designed to clarify the local and varying relations between the terms of a particular description and perhaps this rules out role 5 from consideration. Flutter could also be regarded as the primary topic (role 8). However, firstly, role 8 is not strictly speaking a role at all; it does not show semantic relations between the different terms of an index description, only its subjective state as to a hypothetical reader; it is, in fact, a weighting device. Secondly, it could be argued that piston theory is a primary topic (role 8). Perhaps both terms could be labelled 8, but piston theory could also be regarded as an agent (role 10), although this would overlook the fact that it is not an agent of Flutter but of its analysis. Supersonic flow could be considered as an environment (role 5) or as a cause or influencing factor (role 6). If it is treated as (6) however, Flutter, which it affects, would be the factor influenced (role 7).

When the questions to which document 1014 was one of the relevant documents are considered, the further difficulties in ensuring a match become apparent. Question 97 refers to the Prediction of flutter on lifting surfaces. Flutter could conceivably be given role 3 (undesirable component), or 8 (primary topic). It could also be considered as role 9 in the sense that it is the object of analysis or predictive operations. Two other questions, numbers 98 and 276, to which document 1014 is also relevant, are on how flow characteristics or leading edge bluntness affect Flutter. So now Flutter is an influenced factor (role 7). Yet another question, number 3, is about aeroforces acting on high speed aircraft, and if Flutter is regarded as a kind of aeroforce, it may also be given role 6 (cause or influencing factor).

The general import of these considerations seems to be that the term can be forced to play simultaneously a number of different roles in the same document according to what the particular user is seeking, and that attempting to label them too precisely is liable to result sometimes in unjustifiable rejection of indexing descriptions because they do not match exactly in the roles assigned them.