ELECTRONIC **WORKSHOPS IN COMPUTING**

Series edited by Professor C.J. van Rijsbergen

Jonathan Furner, School of Information and Media Studies, and David Harper, School of Computer and Mathematical Studies, The Robert Gordon University, Aberdeen, Scotland. (Eds)

Information Retrieval Research

Proceedings of the 19th Annual BCS-IRSG Colloquium on IR Research, Aberdeen, Scotland, 8-9 April 1997

Paper:

Automatic Phrase Recognition and Extraction from Text

F. Kelledy and A.F. Smeaton

Published in collaboration with the British Computer Society

BCS

Springer

Automatic Phrase Recognition and Extraction from Text

by

Fergus Kelledy & Alan F. Smeaton,

School of Computer Applications,

Dublin City University,

Glasnevin, Dublin 9, Ireland.

Email: {fkelledy, asmeaton}@compapp.dcu.ie

Abstract: One of the problems facing researchers in the field of Information Retrieval (IR) is that the search criteria used during retrieval (the query) contains terms which are very ambiguous and common. By this we mean that terms can have multiple meanings and occur in a large percentage of the documents in a text collection. Many approaches to addressing this problem have been tried with varying degrees of success. One approach to this problem is to attempt to make the vocabulary used by the IR system less ambiguous by using terms which occur only infrequently. In our case this is achieved through an automatic process of phrase recognition and the incorporation of these phrases into the lexicon of the indexing mechanism used. Unlike previous phrase recognition approaches based on NLP, our work requires no linguistic processing of the text in order to extract phrases but is comparable to what is called 'statistical phrases'. In this paper we describe experiments where we evaluate our phrase recognition on the TREC-4 and TREC-5 collections.

1. Introduction

A major bottleneck in the efficient operation of IR systems is the necessity to process aspects or terms of a query which contribute little in terms of discrimination value and yet consume a large proportion of the processing overhead during retrieval. The standard approach to this problem is to create a stoplist of words **[van Rijsbergen 1979]** which are discarded during the indexing and retrieval processes. This stoplist would include words such as 'and', 'but', 'they', 'which' etc. These words occur very frequently in normal text and as such are not useful in distinguishing one document from another with respect to a given query. Once these stop words have been removed one is left with a restricted vocabulary of words or word stems which should provide an acceptable amount of document discrimination power.

However this is not always the case. Even after stop words have been removed there still exist a large proportion of terms in the lexicon which occur in large percentages of the documents in the text collection, for example the terms '*bank*' and '*computer*' occur frequently within the TREC collection. One could argue that these frequently occurring terms could also be treated as stopwords and discarded from further consideration by the IR system. This in our opinion, is not an acceptable solution to the problem as this would cause the IR system to 'fail' in response to user queries due to users entering queries containing such frequently occurring terms.

While the system's response to such user queries (short queries with commonly occurring terms) may not be very good it is still preferable to the system returning an error stating that it cannot proceed with the query because the terms entered were too general to be contained within the index's lexicon.

One solution to this lexical generality problem is to expand the lexicon rather than restrict it. This expansion involves the recognition of commonly occurring phrases within the text collection and treating these phrases as single entities within the IR system. It has been shown **[Buckley et al 1994]** that phrases are good document discriminators and the use of phrases yield performance improvements in terms of effectiveness. This paper will now detail our approach to identifying phrases with text and how these

phrases are incorporated into the retrieval process, an analysis of the impact of including phrases in the retrieval process in terms of system effectiveness is also presented.

2. Related Research

The idea of using a set of multi-word phrases in addition to single words as a representation for documents and for queries is clearly a desirable feature of any IR systems as phrases are normally more content-bearing and meaningful. Automatically identifying phrases in text can be broadly classified into linguistically-based and statistically-based techniques and this division has been around for at least 10 years.

If we use the recent TREC-5 exercise **[TREC-5 1996]** as an indicator of phrasal indexing techniques, we find a few examples of systems which index documents and queries by identifying phrases from text based on simple word occurrence and co-occurrence patterns. The system developed at IBM is typical of this as is the work from Cornell and from George Mason University. As an alternative to identifying phrases at index time and using them as entries in the database's lexicon, other approaches allow a user to specify a phrase in their query and then at retrieval time the co-occurrence or adjacency of query phrase components is rewarded in a document's score. This is the approach taken in many web search engines and it simulates the technique of indexing documents by phrases without actually having to do this or causing any overhead in the inverted file. In general these approaches tend to be lightweight in terms of computational overhead though there is an extra cost at retrieval time to combine inverted file entries to identify query phrase occurrences.

Linguistically-based approaches to recognising phrases would seem to be intuitively more desirable than statistical as true phrases can be identified in text rather than statistical co-occurrences. One of the first such approaches was taken in the FASIT system **[Dillon & Gray 1983]**. In this system, word tokens in document texts were assigned one or more likely part-of-speech tags and a set of rules which searched for POS tag patterns known to indicate content-bearing terms was used. Fagan **[Fagan 1987]** was one of the first to apply a syntactic analysis to the full text of documents and queries and use this to derive 2- and 3-word phrases based on syntactic constructs. Fagan also used a simpler statistically-based technique to identify phrases and in a direct comparison between the two approaches he found not much difference between the two in terms of retrieval effectiveness though the statistical approach was more desirable because of its much lighter computational processing costs.

A full parse of c. 2 Gbytes of document texts in the TREC exercise has been performed by the group from General Electric led by Tomek Strzalkowski. From this parse, various phrase recognition approaches have been tried, for example identifying dependent word pairs like heads of clauses and their modifiers and using these as the representation of documents. This has led to improvement in retrieval effectiveness but at an enormous cost in computational overhead due to the linguistic analysis.

At the University of Manitoba in Canada, a group has been participating in recent MUC conferences **[Lin 1995]**, similar to TREC but targeted at information extraction rather than information retrieval. The approach taken here is to use a set of rules for identifying linguistic phrases such as compound nouns and adjective/noun collocations and use these for representing text content. In the MUC framework this has performed quite well but MUC is small-scale compared to TREC and the technique has yet to be scaled up.

Once a set of phrases to represent document content has been identified then the retrieval operation can default to the term weighting and document ranking techniques used for single-word representations and this is what is generally done in most of the work described above. A huge problem which remains however, is normalising phrases. It is a feature of natural language that there are many syntactically different ways of saying the same thing ("a red-coloured empty petrol can" is the same as "an empty red can of petrol". Identifying these as equivalent cannot reply on simply counting the number of words in common as the examples "blind venetians" vs. "venetian blinds" and "juvenile victims of crime" vs. "victims of juvenile crime", show.

The issues of phrase matching were highlighted in **[Smeaton 1992]**, showing the approaches to this being to ignore, to allow some kind of graded match between phrases, or normalise phrases into one

standardised format. The latter approach is taken by the CLARIT system which avoids the phrase matching problem by using a linguistic analysis of a document sample to define the vocabulary to be used in the lexicon of the system. All documents are then processed by a syntactic analysis of texts which maps phrases occurring in the document into this restricted pre-defined vocabulary. This approach has been shown, repeatedly, to yield an effective document and query representation. The MUC system developed by the University of Manitoba **[Lin 1995]** takes a different approach by indexing a text into phrases and then matching normalised dependency trees derived from these phrases where the normalisation handles syntactic variations. This is similar to an approach taken by the authors in an earlier TREC **[Smeaton *et al* 1992]**. In summarising related work on phrase indexing we must note that the true effect on retrieval quality of using phrases rather than single words to index text tends to be hidden in reported work because there are normally so many other "ingredients" used in IR experiments which prevent a direct comparison of using phrases vs. not using them. In TREC-5, the Xerox group reported that terms derived from a shallow parse of texts performs marginally better than using simple contiguous 2-word units but the differences are not major but the question of statistically-based vs. linguistically-based is still an open one. Of the techniques which have been reported, they vary from the very simple to the very elaborate yet are all comparable, as far as can be judged.

3. Motivation for Research

Our retrieval system was primarily developed to investigate efficiency / effectiveness trade-off issues in IR. The main focus of our research is the development of techniques which facilitate the efficient processing of large automatically expanded user queries through a process of restricting the amount of information processed during retrieval. In order to provide more effective retrieval we decided to incorporate the use of phrases in our system and in order to do this we needed a set of phrases for matching purposes.

The purpose of this research was to determine the possible benefits of using a statistically defined phrase vocabulary in an IR system. In order to define a baseline for our experiments we needed a pre-defined phrase vocabulary. This we got from WordNet[1], by simply extracting all of the phrases from the WordNet lexicon. This provided us with a predefined set of around 35,000 valid phrases to match against the documents in the text collection. After creating the index using the above set of WordNet phrases we found that only 5% of the terms in the lexicon were phrases and the rest were all single terms. This had the effect of only having a very small subset of the total number of query terms used for evaluation of the system treated as phrases. These initial results confirmed our expectations that the phrase vocabulary must be generated from the document set itself as in CLARIT[2] and not derived from an independent third party source. We then set out to develop a method which would automatically extract phrases from the documents being indexed by the IR system.

4. Phrase Recognition

In order to extract phrases from text we must have a clear idea what exactly qualifies as a phrase. Within our test environment we regard any commonly occurring sequence of terms as a possible phrase. There are a number of restrictions to this rule. Firstly the phrase cannot begin with a stopword. This means that phrases such as '*the car*' and '*a house*' are not valid phrases. Secondly, phrases cannot end with a stopword. This eliminates phrases such as '*buy a*' and '*play the*'. Thirdly, phrases can contain stop words and this allows phrases such as '*department of defence*' and '*sitting on the fence*'. Another criteria for a term's inclusion into a phrase is the occurrence of consecutive terms all of which start with capital letters. This criteria is included to aid the recognition of commonly occurring names of people e.g. '*George Bush*' , company names, e.g. '*International Business Machines*', and place names, e.g., '*New York*' and '*San*

[1] WordNet is a semantic knowledge base developed by Princeton University.

[2] CLARIT is an Information Retrieval system developed by CLARITECH.

Francisco'. We must also formally define what constitutes 'commonly occurring' phrases, i.e. how frequently term co-occurrences must occur in order for them to be classified as phrases.

Due to the way our IR system had been developed, the phrase recognition procedure was easily implemented by taking an existing document pre-processing module of our IR system and modifying it. The easiest way to explain its operation is by example. Take the following extract from the TREC[3] text collection:

> *The celluloid torch has been passed to a **new generation**: filmmakers who grew up in the 1960s. ``Platoon,'' ``**Running on Empty**,'' ``1969'' and ``**Mississippi Burning**'' are among the movies released in the past two years from writers and directors who brought their own experiences of that turbulent decade to the screen. ``The contemporaries of the '60s are some of the filmmakers of the '80s. It's natural,'' said **Robert Friedman**, the senior vice president of world wide advertising and publicity at Warner Bros. **Chris Gerolmo**, who wrote the screenplay for ``**Mississippi Burning**,'' noted that the sheer **passage of time** has allowed him and others to express their feelings about the decade. ``Distance is important,'' he said. ``I believe there's a lot of thinking about that time and America in general.'' The **Vietnam War** was a defining experience for many people in the '60s, shattering the consensus that the **United States** had a right, even a **moral duty** to intervene in conflicts **around the world**. Even today, politicians talk disparagingly of the ``**Vietnam Syndrome**'' in referring to the country's reluctance to use **military force** to **settle disputes**.*

The highlighted portions of the text extract are possible candidates for classification as phrases, they include frequently occurring phrases such as '*new generation*', '*moral duty*' and '*around the world*', and person names '*Robert Friedman*' and '*Chris Gerolmo*'. Word co-occurrences like '*celluloid torch*' would not be treated as a phrase due to its relatively infrequent occurrence in the overall document collection.

It must be remembered that meaningful phrases cannot be extracted just from this extract of text alone. The phrase recognition process only becomes effective when a large amount of textual information is processed. This allows statistical information to be gathered on the frequency of occurrence of phrases. Figure 1 illustrates the method used to extract phrases from the documents text:

W1	W2	W3	Output
been	passed	to	None
passed	to	a	None
to	a	new	None
a	new	generation	None
new	**generation**	filmmakers	Yes

Figure 1 - Example sliding text window.

This sliding window (see Figure 1), which for this set of experiments is limited to three words, moves through the documents text when the window begins with a stop word then that word is skipped and the contents of the window are shifted left by one position. If the first word is a non stopword then output is produced only if the ending term is not a stopword. Once a candidate phrase has been located it is stored, if this is the first occurrence of the candidate phrase then a new storage structure is allocated to it and the candidate phrase's document identifier is also stored.

Once all of the documents have been processed the output is a list of candidate phrases and all of the documents they occur in. This list is then sorted by phrase name and a count of the number of unique document identifiers in which the phrase occurs in is taken. It is this count which determines whether or not the candidate phrase is included into the phrase set for the collection.

In our implementation the recognition process involves processing the text in sections, with each section being around 20 Mbytes. This section size is determined by the amount of memory available for the process on the machine. For each section of text, a file containing each possible candidate phrase along

---

[3] TREC is the Text Retrieval Conference run by NIST, the National Institute of Standards and Technology in Washington DC since 1992.

with its occurrence frequency is produced. This file is then sorted by decreasing occurrence frequency and all phrases with an occurrence frequency of greater than 25 (i.e. the phrase occurs in more than 25 different documents within the current text section) is included in the phrase set. The selected phrases are then added to a global phrase set.

The phrase extraction process is a once off event for static text collections and a periodic one for dynamic text collections. The stopping criterion for the phrase extraction procedure is as follows: the process is repeated until the number of new phrases being added to the global phrase set falls below a certain threshold value. At this point it can be assumed that the vast majority of the phrases have been located and extracted. This assumption depends on the text collection being static in nature. If the IR system was dealing with a dynamic text collection then the phrase recognition procedure would have to be run periodically when the amount of new documents added to the collection allowed the statistical extraction of new phrases.

5. Characteristics of Phrases

The phrases generated from this approach have a distinct advantage over the pre-defined set of phrases extracted from WordNet in that they are extracted from the text collection being indexed therefore the number of phrase matches obtained using these phrases will be much higher. This however does not limit their use to this text collection only. Once generated the phrases can be used as a base phrase set which can augment phrases generated from other text collections. It must also be noted that all phrases generated by this process are stemmed, this means that phrases like '*computer department*' and '*computing departments*' will be reduced to their base form '*comput depart*'. This has the effect of increasing the probability of matches between phrases and terms within the documents. At present we are limiting the length of the phrase to three terms. The pre-defined phrase set from WordNet contained phrases of up to six terms but an analysis of the frequency of occurrence of these longer phrases showed that they occurred very infrequently.

Once the global phrase set has been compiled it is then incorporated into the document pre-processing procedure of the indexing system. The same sliding text window approach described above is applied to the documents. Each word entering this window is stemmed. The stemmed words are checked to see if they are the possible beginning of a phrase. If so then the rest of the words in the current text window are matched against the global phrase set to see if a match occurs. It is possible for more than one match to occur for example the sentence '*President Bush said the federal budget deficit is holding steady*', will extract the phrases '*federal budget*' and '*federal budget deficit*'. It must be noted that the individual phrase component terms will also be included in the documents internal representation within the system.

The internal representation of the above sentence is as follows: '*President Bush*', '*President*', '*Bush*', '*federal budget deficit*', '*federal budget*', '*federal*', '*budget*', '*deficit*', '*holding*', '*steady*'. The stemmed version of these terms and phrases are then used to index the sentence and queries containing any of these terms or phrases will achieve a positive match with this sentence.

6. Implications of using Phrases in IR Systems

The inclusion of phrase recognition improves the precision of the retrieval process. For example, given a query containing the terms '*Bill Gates*', if no phrase recognition process is employed then this query will be treated as two separate terms '*Bill*' and '*Gates*'. These terms will be stemmed to '*Bill*' and '*Gate*'. As such, documents containing the term '*Bill*' and documents containing '*Gate*' will be returned. A document containing the sentence '*Bill and John opened the gate*' will match the query and achieve a reasonable query document similarity score, but it is obvious even from this simple example that the document is not what the user was looking for. Now if phrase recognition is employed the query '*Bill Gates*' becomes '*Bill*', '*Gates*', '*Bill Gates*'. These query terms and phrases are stemmed to '*Bill*', '*Gate*' and '*Bill Gate*'. The document containing the sentence '*Bill and John opened the gate*' will still achieve a reasonable query document similarity score, however the document containing the sentence '*Bill Gates, the founder of Microsoft*', will match on all three components of the query, namely the two query terms and the query phrase. If standard term weighting strategies are employed such as TF*IDF weighting then occurrence

frequency of the phrase '*Bill Gate*' will be less than the occurrence frequencies of the terms '*Bill*' and '*Gate*' and hence it will get an higher inverse document frequency (IDF) score. This will have the effect of increasing the query-document similarity score of the document containing the sentence '*Bill Gates, the founder of Microsoft*' returning it to the user before the first document.

It is clear that the use of phrases in an IR system could have a beneficial effect in terms of efficiency as the inclusion of phrases into the indexing and retrieval procedures alters the structure of the index from an index with a relatively small lexicon (small number of unique terms) and a large average posting list length (most postings occur frequently in text collection) to an index with a large lexicon (terms and phrases) with a shorter average posting list length. The decrease in the average posting list length is achieved by the incorporation of the phrases into the lexicon which tend to occur less frequently than individual and hence have shorter posting lists. If a large enough proportion of the index terms activated during retrieval are phrases then the processing of the query is altered by the need to handle a larger number of relatively short posting lists.

7. Results

In order to evaluate the impact of including phrases in the IR system's lexicon we tested the system with no phrases, with phrases derived from WordNet, with automatically generated phrases (from the text being indexed) and with a combination of WordNet and automatically generated phrases. The phrases derived from WordNet exhibited the following characteristics:

Phrase Length (in terms)	Occurrence Frequency
7	1
6	9
5	57
4	416
3	3,762
2	31,142

Figure 2 - Statistics of WordNet phrases.

As illustrated in Figure 2 the vast majority of phrases are composed of three or less component terms. This fact lead us to restrict the phrase length in our automatic phrase extraction procedure to three or less component terms. These phrases were generated from processing disks 1 & 2 of the TREC collection (see Figure 3).

Phrase Length (in terms)	Occurrence Frequency
3	65,157
2	120,351

Figure 3 - Statistics of automatically generated phrases.

These automatically generated phrases are derived from the text being indexed and therefore should have a higher probability of being of use during retrieval by discriminating effectively between document in the collection. Figure 4 illustrates the statistics of the combined phrase set (WordNet and automatically generated).

Phrase Length (in terms)	Occurrence Frequency	Overlap
7	1	0
6	9	0
5	57	0
4	416	0
3	68,846	73
2	150,441	1,051

Figure 4 - Statistics of merged phrase set.

Figure 4 illustrates the extremely low degree of overlap between the WordNet phrases and the automatically generated phrases with only 0.1% of the three-term WordNet phrases and 0.69% of the two term WordNet phrases being automatically generated by our phrase extraction approach. This would suggest that the WordNet phrase collection and for that matter any independently generated phrase collection would not be as effective as a phrase collection generated from the text being indexed simply because the degree of overlap between the independent phrases and the text collection would be too low to allow a significant number of phrases to be incorporated into the indexing and retrieval processes.

In order to test the impact of the inclusion of automatically generated phrases on our retrieval performance we carried out the following experiments. We tested the TREC-4 and TREC-5 collections with no phrases and with the automatically generated phrases. The results for these four experimental runs are summarised in Figure 5.

	TREC-4 No Phrases	**TREC-4 Phrases**	**TREC-5 No Phrases**	**TREC-5 Phrases**
Retrieved:	50000	50000	50000	50000
Relevant:	6503	6503	5524	5524
Returned:	2707	2685	2054	2096
at 0.00	0.4704	0.5023	0.3798	0.4094
at 0.10	0.2350	0.2874	0.2072	0.2523
at 0.20	0.2005	0.2247	0.1658	0.2012
at 0.30	0.1638	0.1881	0.1326	0.1638
at 0.40	0.1293	0.1517	0.1171	0.1387
at 0.50	0.0954	0.1178	0.0911	0.1132
at 0.60	0.0678	0.0893	0.0707	0.0864
at 0.70	0.0411	0.0615	0.0471	0.0671
at 0.80	0.0293	0.0430	0.0258	0.0393
at 0.90	0.0014	0.0107	0.0102	0.0224
at 1.00	0.0014	0.0014	0.0032	0.0062
Average Precision:	**0.1084**	**0.1325**	**0.0961**	**0.1172**
P @ 5 docs	0.2440	0.3000	0.1880	0.2280
P @ 10 docs	0.1960	0.2500	0.1900	0.2240
P @ 30 docs	0.1620	0.1947	0.1540	0.1813

Figure 5 - Precision-Recall Figures for No Phrases versus Phrases for the TREC-5 collection.[4]

Figure 6 illustrates the effect of including phrase recognition in the retrieval process for the TREC-4 and TREC-5 query sets. The number of unique terms recognised increases by 68 terms for the TREC-4 collection and by 31 terms for the TREC-5. These extra terms represent phrases recognised by the system within the queries. These extra phrases are the only difference between the two TREC-4 runs and the two TREC-5 runs and as such the improvement in results in terms of Precision-Recall can be attributed to the inclusion of these extra phrases in the two query sets.

	TREC-4	TREC-5
Total No. unique terms (ex. Phrases)	352	1465
Total No. unique terms (incl. Phrases)	420	1496

Figure 6 - Effect of Phrase inclusion on the TREC-4 and TREC-5 query sets.

8. Conclusions

It is clear that the use of phrase recognition in whatever form improves the precision of an IR system by providing a much richer vocabulary in the index's lexicon. This enriched vocabulary is less ambiguous

[4] Precision-Recall figures generated using the 'standard' TREC evaluation routine.

(due to the fact that phrases tend to be less ambiguous than individual terms) and occurrence frequency of the phrase is normally much lower than the individual terms thus increasing their inverse document frequency scores therefore providing more discriminating terms in the index's lexicon.

The contribution of the work reported here is that we present another phrase-recognition algorithm to determine the lexicon used for indexing a corpus of text. A moot point in IR research as to the necessity of having to involve a computationally demanding NLP analysis technique as is done for example in CLARIT and the work of Strzalkowski versus the much more lightweight approach of identifying 'statistical phrases'. Although intuitively less appealing than using NLP techniques, the simple technique such as the one we have presented here continue to perform well and show improvement over using single terms alone (as illustrated in Section 6). Our particular phrase recognition strategy which is carried out as a pre-parse of the collection being indexed is relatively undemanding in terms of resources required. In addition, our phrase recognition process can be applied to languages other than English once a stoplist and stemmer is available and we have used it as part of our indexing of Spanish documents and queries in TREC-5.

From our initial experiments this automatic generation of phrases from text has shown promising results. Resulting from the experience gained from these experiments it would be interesting to determine whether or not there is an optimum ratio of terms to phrases in the index's lexicon. Another avenue of research would be to determine whether there is an optimal phrase length, i.e. is it enough to restrict phrase length to two terms or are there additional improvements to be gained by incorporating longer phrases into the index's lexicon.

Another research possibility in this field is to determine whether the generation of lexical variants of the automatically generated phrases would be of benefit. By this we mean deriving phrases like '*defence department*' from the automatically generated '*department of defence*'. The system at present will not get a match between these two phrases when perhaps it would be of benefit.

9. References

[Buckley et al 1994] C. Buckley, G. Salton, J. Allan, A. Singhal, 'Automatic Query Expansion Using SMART : TREC 3', TREC 3, National Institute of Standards and Technology, Washington DC, USA, November 1994.

[Dillon & Gray 1983] M. Dillon, A. S. Gray, 'FASIT: a fully automatic syntactically-based indexing system', Journal of the ASIS, Vol. 34, No. 2, pages 99-108, 1983.

[Fagan 1987] J. L. Fagan, 'Automatic phrase indexing for document retrieval: an examination of syntactic and non-syntactic methods', Proceedings of the Tenth ACM SIGIR Conference on Research and Development in Information Retrieval, pages 91-108, June 1987.

[Lin 1995] D. Lin, 'Description of the PIE System as Used for MUC-6', Proceedings of the Sixth Conference on Message Understanding (MUC-6), Columbia, Maryland.

[Smeaton 1992] A. F. Smeaton, 'Progress in the Application of Natural Language Processing to Information Retrieval Tasks', The Computer Journal, Vol. 35, No. 3, 1992.

[Smeaton et al 1992] A. F. Smeaton, R O'Donnell, F. Kelledy, 'Indexing Structures Derived from Syntax in TREC-3: System Description', Proceedings of TREC-3, NIST Special Publication 500-225, 1995.

[TREC-5 1996] Proceedings of the TREC-5 conference, National Institute of Standards and Technology, Washington DC, USA, November 1996.

[van Rijsbergen 1979] C. J. van Rijsbergen, 'Information Retrieval', Second Edition, Butterworths & Co. Ltd., 1979.