Are Titles of Chemical Papers Becoming More Informative?

The efficiency of key-word-in-context (KWIC) permuted-title indexes and their numerous variations is highly dependent upon authors' choices of titles for their papers. Titles are important not only in commercial services, such as *Chemical Titles, BASIC, Current Contents,* and *CA Condensates,* but also in scanning primary journals, and in traditional library services, such as bibliographies. It is generally believed and often stated that titles of chemical papers are becoming more informative as authors become increasingly aware of the importance of titles as "carriers" of information. The present study was undertaken to test whether (1) titles of chemical papers are becoming more informative and (2) whether uninformative titles of chemical papers are being eliminated since the advent of the KWIC index in 1958.

The first hypothesis was tested by comparing titles published in 1948, 1958, and 1968 by the following criteria: (1) a count of substantive words in the title; (2) a count of all word matches between title and abstract, with and without the use of a thesaurus; and (3) a count of word matches between title and 10 leading substantive words selected from the abstract, with and without the use of a thesaurus. The second hypothesis was tested by comparing a count of short titles (with 3 or less substantive words) published in 1948, 1958, and 1968.

Results confirm that uninformative titles of chemical papers are being eliminated and that informative titles are becoming more informative since the advent of the KWIC index.

JACQUES J. TOCATLIAN †

Merck Sharp & Dohme Research Laboratories
West Point, Pennsylvania

• Introduction

The key-word-in-context (KWIC) permuted-title index was introduced by H. P. Luhn in 1958 as a prompt, relatively inexpensive means of building a temporary bridge between the contents of the current literature and readers, awaiting the completion of the more slowly prepared conventional indexes. Luhn himself recognized that the quality of the KWIC index could not equal that of some of the more carefully prepared conventional indexes. However, because of its low cost and ease of preparation, the KWIC index and its variations have replaced the conventional index in many cases.

One of the strongest objections voiced against such indexes is that titles of papers are unsatisfactory as a basis for subject index entries because titles are not composed with indexing in mind. The adequacy of titles as a source of subject content clues has been given much attention in the last decade, and many studies have been published on the subject.

The results and conclusions of these studies on the adequacy of titles vary. There is, however, a recognition of the obvious: The efficiency of a KWIC index is dependent upon authors' choices of titles. In fact, the importance of informative titles has been stressed in the literature for many years.

There developed a frequently stated conviction that the very existence of permuted-title indexes would stimulate authors to use better titles for their papers, as exemplified by J. D. Black's statement in 1962: "Before long the engineer, scientist, or mathematician will realize that if his title is not descriptive enough his paper will not be used as much as it might be."

The belief that titles would become more informative is sometimes translated into statements that they actually did. However, none of the statements made verbally at professional meetings or in writing were based on statistical evidence.

Recognition that evidence in support of, or in contradiction to, the belief that titles are becoming more informative could be important in future index building led to the undertaking of the present study.

† Present address: UNESCO, Paris, France.

Titles are important as "carriers" of information not only in permuted-title indexes, such as *Chemical Titles* and *BASIC* (the index to *Biological Abstracts*), but also in tape services, such as *CA Condensates*, and in widely read alerting services, such as *Current Contents*. Titles are important for readers who scan the primary journals, and in traditional library services, such as bibliographies and lists of references.

It seemed important to ask at this point, Are titles of chemical papers really becoming more informative?

The present study tests the following two hypotheses:

1. As rated by the criteria adopted and validated, titles of chemical papers are becoming more informative since the advent of the KWIC index.
2. As rated by the criterion adopted, uninformative titles of chemical papers are being eliminated since the advent of the KWIC index.

Both hypotheses play a part in how informative titles are: Elimination of uninformative titles results in a proportional increase in informative ones; independently, informative titles may become more informative.

• Hypothesis No. 1

The first hypothesis (titles are becoming more informative) was tested by using 5 criteria, or measurements, to compare titles of chemical papers published in 1948, 1958, and 1968.

CRITERIA

Measurement A: *all substantive words*—a count of all substantive words in the title.

For the purpose of this work it was found practical to isolate substantive words by rejecting all nonsubstantive words. These were defined as noninformative words that convey little or no information by themselves, such as articles, prepositions, conjunctions, pronouns, and auxiliary verbs. An example will be given shortly.

Measurement B: *all matches*—a count of the number of substantive words in the title that match corresponding words in the abstract, using a thesaurus concept in making matches.

The thesaurus allowed a match between the singular form of a word and its plural form; between synonyms; between specific and generic terms; between a chemical compound and its formula; between verbs in different tenses; and, in general, between terms conveying the same concept. Multiple appearances of key words in titles were counted as one appearance.

Measurement C: *level matches*—a count of the number of substantive words in the title that match identical substantive words in the abstract, without using a thesaurus.

In this case nouns in singular and plural forms or verbs in different tenses were not counted as matches; neither were synonyms nor chemical compounds and their formulas.

The difference between measurements *B* and *C* is the use in *B* of a thesaurus in making matches.

Before subjecting a title to measurements *D* and *E*, each abstract was read and 10 substantive words that seemed to describe best the subject content of the abstract were selected (referred to as 10 "leading" terms). The choice was subjective, in much the same way an indexer selects indexing terms to describe a paper.

Measurement D: *all leading matches*—a count of the number of substantive words in the title that match any of the 10 "leading" terms, using a thesaurus.

Measurement E: *level leading matches*—a count of the number of substantive words in the title that match any of the 10 "leading" terms, without using a thesaurus.

The results of the following title subjected to measurements *A* through *E* are shown in Table 1.

TABLE 1. Results of measurements A through E made on a sample title

A All substantive words	*B* All matches	*C* All level matches	10 leading terms	*D* All leading matches	*E* Level leading matches
$A = 6$	$B = 5$	$C = 2$		$D = 4$	$E = 2$
Field	Field	Field	Analysis	—	—
Method	Test	—	Test	Method	—
Rapid	—	—	Color	—	—
Determination	Determining	—	Determination	Determination	Determination
Hydrogen cyanide	Hydrogen cyanide	Hydrogen cyanide	Hydrogen cyanide	Hydrogen cyanide	Hydrogen cyanide
Air	Atmospheres	—	Atmospheres	Air	—
			Industrial	—	—
			Paper	—	—
			Ferrous sulfate	—	—
			Sodium hydroxide	—	—

A FIELD METHOD FOR THE RAPID DETERMINATION OF HYDROGEN CYANIDE IN AIR

B. E. Dixon, G. C. Hands and A. F. F. Bartlett
(Department of the Government Chemist, Clement's Inn Passage, Strand, W.C.2)

A field test for determining small amounts of hydrogen cyanide in industrial atmospheres is based on formation of a Prussian blue colour on test paper impregnated with ferrous sulphate and sodium hydroxide. The test is specific for hydrogen cyanide. The behaviour of a number of possibly interfering gases has been investigated. The test is sensitive to slightly less than 1 p.p.m. of hydrogen cyanide in air and has an error of ± 10 to 20 per cent. The blue stains obtained are permanent. Test papers properly prepared and stored retain their activity for at least 10 months.

Analyst, 83: 199 (1958)

The following terms, shown in Table 1, were considered equivalent:

determination = determining
air = atmosphere
method = test

Selection of Sample

The criteria for journals to be included were that they be:

1. In the field of chemistry
2. In the English language
3. Published continuously since 1948
4. Committed to publishing an abstract or summary with every article

The universe selected for this study was confined to the chemical journal collection of an Industrial Research Library (Merck & Co., Inc.; Rahway, N.J.). From the periodical holding list of the 736 journals in this library, it was found that 96 were in English with issues back to at least 1948. Of these, 64 carried an abstract with each article. Ten out of the 64 dealt with chemistry and, therefore, constituted the journal sample for this study.

The following journals fulfill the selection criteria and are complete for the years 1948, 1958, and 1968, except as noted in parentheses:

1. *Acta Chemica Scandinavica* (1968: missing August through December)
2. *Analyst* (1968: missing December)
3. *Analytical Chemistry*
4. *Discussions of the Faraday Society* (1967: complete)
5. *Journal of the American Chemical Society*
6. *Journal of the Chemical Society* (1968: missing B11, C22, and C25)
7. *Journal of Organic Chemistry*
8. *Journal of Physical Chemistry*
9. *Journal of Research of the National Bureau of Standards*
10. *Transactions of the Faraday Society* (1968: missing January through June and December)

Using a table of random numbers, 10 articles for each of the 3 years for each of the 10 journals were selected. The total number of articles per journal was, therefore, 30, and the total number of articles in the sample was 300.

Validation of Measurements A Through E

Before any measurement was taken, the 300 titles constituting the sample were read and a subjective selection was made of 20 informative titles and 20 uninformative titles. The articles themselves, or their abstracts, were not read; the subjective choice was made based exclusively on titles. No attempt was made to select the 20 best, or *most* informative titles, and the 20 *least* informative; but, rather, the first 20 titles encountered that seemed to convey a fair amount of information about the article and, conversely, the first 20 titles that seemed too general, vague, ambiguous, or uninformative were chosen. These titles were used for the validation of measurements; some examples follow.

Informative Titles

1. Dinickel Phosphide as a Heterogeneous Catalyst for the Vapor Phase Reduction of Nitrobenzene With Hydrogen to Aniline and Water (*Journal of the American Chemical Society*)
2. The Heat Capacity, Heat of Fusion and Entropy of Benzene (*Journal of the American Chemical Society*)
3. The Kinetics of Nitrogen Formation From Nitrous Acid and Ammonium or Methylammonium Ions (*Journal of Physical Chemistry*)
4. Vulcanization of Synthetic Rubbers by the Peachy Process (*Journal of Research of the National Bureau of Standards*)
5. The Determination of Magnesium in Silicate and Carbonate Rocks by the Titan Yellow Spectrophotometric Method (*Analyst*)

Uninformative Titles

1. A Note on the Chemistry of Enterogastrone (*Acta Chemica Scandinavica*)
2. Variability and Inhomogeneity of Aluminum Dilaurate (*Journal of Physical Chemistry*)
3. Some Physical Properties of Uranium Hexafluoride (*Transactions of the Faraday Society*)
4. Studies on the Chemistry of Lichens (*Acta Chemica Scandinavica*)
5. The Preparation of Some Iodinated Phenylalkanoic Acids (*Journal of Organic Chemistry*)

Table 2 shows the results obtained when these 20 informative and 20 uninformative titles were subjected to measurements *A* through *E*.

Table 2. Measurements A through E for informative and uninformative titles

Type of title	*A* All substantive words	*B* All matches	*C* Level matches	*D* All leading matches	*E* Level leading matches
Mean value of 20 informative titles	5.8	5.3	5.3	5.0	4.6
Mean value of 20 uninformative titles	4.4	2.2	1.7	2.1	1.6
Difference between informative and uninformative titles	1.4	3.1	3.6	2.9	3.0

The standard analysis of variance technique was used for all measurements. The differences in measurements A through E between the 20 informative and 20 uninformative titles are statistically significant (probability less than 0.001). These results support the validity of the criteria adopted for the analysis of titles.

It might be suggested that these criteria be used as a general method for evaluating titles. It seems that this general method could very well be applied to the analysis of titles in other disciplines, such as Biology and Engineering, and the results compared to those presented here.

Results

Table 3 shows the mean values of measurements A through E for the 10 journals. In all cases no statistically significant difference was detected between 1948 and 1958, whereas, a statistically significant difference was found between 1958 and 1968 for all measurements.

Table 4 shows the difference between decades in mean values of A through E. We can see that titles of chemical papers in 1968 contained, on the average, a statistically significant increase in the number of substantive words over those of 1958 or 1948 and a statistically significant increase in the number of substantive words that matched corresponding words in the abstract over those of 1958 or 1948. When substantive words in the title were compared with 10 leading terms selected subjectively from the abstract to describe its content, there was a statistically significant increase in the number of matches in 1968 chemical papers over those of 1958 or 1948.

Figure 1 shows a plot of the mean values of measurements A through E versus time. We can, therefore, conclude that our first hypothesis is valid: *As rated by the criteria adopted, titles of chemical papers are becoming more informative since the advent of the KWIC index.*

• Hypothesis No. 2

To test the second hypothesis (uninformative titles of chemical papers are being eliminated since the advent of the KWIC index), we compared a count of short titles (with three or less substantive words) published in 1948, 1958, and 1968. The following is a list of examples of titles with three or less substantive words:

1. New Hafnium Phosphides (*Acta Chemica Scandinavica*, 1968)
2. 3,3-Dinitro-1-Alkanols (*Journal of Organic Chemistry*, 1958)
3. Slow Combustion of Methyl Ethyl Ketone (*Transactions of the Faraday Society*, 1968)
4. Heats of Adsorption (*Journal of Physical Chemistry*, 1948)
5. Algenic Acid Acetate (*Journal of the Chemical Society*, 1948)
6. Recent Studies on Sialic Acid (*Acta Chemica Scandinavica*, 1958)
7. The Viscosity of Dilute Emulsions (*Transactions of the Faraday Society*, 1958)
8. Some N-Arylsulfonyl-N'-Alkylureas (*Journal of Organic Chemistry*, 1958)

Table 3. Mean values of measurements A through E for the 10 journals

[100 titles per year]

Measurement	1968	1958	1948
A: All substantive words	6.77	5.46	5.57
B: All matches	4.83	4.02	4.13
C: Level matches	4.17	3.39	3.64
D: All leading matches	4.35	3.65	3.62
E: Level leading matches	3.67	3.01	3.08

Table 4. Difference between decades in mean values of A through E

[Mean values are given for 100 titles per year]

Difference	*A* All substantive words	*B* All matches	*C* Level matches	*D* All leading matches	*E* Level leading matches
1968–1958	1.31	0.81	0.78	0.70	0.66
1968–1948	1.20	0.70	0.53	0.73	0.59
1958–1948	—0.11	—0.11	—0.25	0.03	—0.07

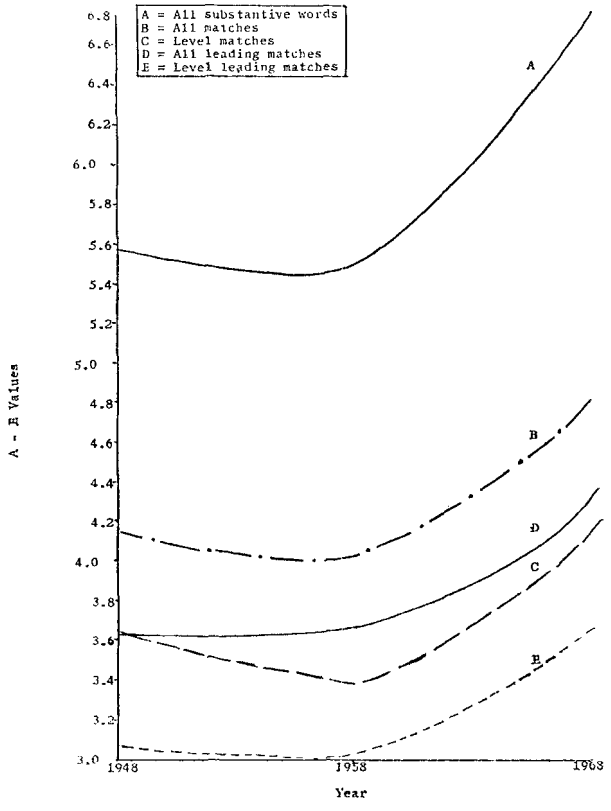


Fig. 1. Mean values A through E versus time

9. Phase Equilibrium Description (*Journal of Physical Chemistry*, 1948)
10. Determination of Aspartic Acid (*Analytical Chemistry*, 1948)
11. Diffusion Studies With Lysolecithin (*Journal of the Chemical Society*, 1958)
12. Synthesis of Ketodextrans (*Acta Chemica Scandinavica*, 1968)
13. Octaethylporphyrin (*Journal of Organic Chemistry*, 1968)
14. Microeffusiometry (*Analytical Chemistry*, 1948)
15. Investigations in Serum Copper (*Acta Chemica Scandinavica*, 1948)

How many substantive words are needed to make an informative title?

It would be very difficult, if not impossible, to state the answer quantitatively. However, "six substantive words" has been stated arbitrarily in the literature as a minimum number to make an informative title. *Chemical Titles*, for example, had, in 1963, an average of six key words per title.

In the present study, a sample of 20 titles selected subjectively as being informative showed an average of 5.8 substantive words per title. On the other hand, a sample of 20 titles selected subjectively as being uninformative showed an average of 4.7 substantive words per title.

It seemed logical to assume that titles with three or less substantive words are most likely uninformative, the thinking being that if six substantive words represent an average, four to eight could be taken as the range.

It was reasoned that by counting the number of titles with three or less substantive words in a given journal for 1968, 1958, and 1948, and comparing the percentages of such titles, it could be established whether there has been a discernible tendency for the uninformative title to disappear. If uninformative titles are being eliminated, this is one means by which titles are becoming, overall, more informative.

By this criterion, the title of this article would not rate as informative. In fact, it is not a good title; it asks a question but does not indicate what was done.

Selection of Sample

Every article published in 1948, 1958, and 1968 for the 10 journals selected for the study was taken into account. The only exception was the articles in the issues missing from the collection at the time of sampling and the articles in French or German published in *Acta Chemica Scandinavica*.

Number of articles for 1948	3,061
Number of articles for 1958	5,780
Number of articles for 1968	7,499
Total number of articles 1948–68	16,340

Results

By observing the data obtained (Table 5), it was obvious that journals fell in three general groups. The first group (*Journal of Organic Chemistry*) shows 31.7% titles with three or less substantive words in 1948, decreasing to 29.6% in 1958, and dropping steeply to 15.9% in 1968. It would seem that any improvement observed in the titles of this first group from 1948 to 1968 would derive from the elimination of short titles (T_{123}). Measurements A through E for this group did not detect an improvement in the titles (statistically significant difference in any A through E value over the years).

The second group consists of the three fundamental journals *Acta Chemica Scandinavica, Journal of the American Chemical Society,* and *Journal of the Chemical Society*. The mean value of T_{123} for this group decreased from 24.5% in 1948, to 13.3% in 1958, to 10.9% in 1968. This group of journals had a lower percentage of short titles for the years 1948, 1958, and 1968 than group I.

Table 5. Mean values (90% confidence limits) of the number of titles with three or less substantive words

Journal	Mean value of T_{123}		
	1968	1958	1948
Group I (*Journal of Organic Chemistry*)	15.9% (14.24%, 17.64%)	29.6% (27.02%, 32.25%)	31.7% (25.54%, 38.22%)
Group II (*Acta Chemica Scandinavica, Journal of the American Chemical Society,* and *Journal of the Chemical Society*)	10.9% (10.13%, 11.70%)	13.3% (12.34%, 14.29%)	24.5% (22.94%, 26.10%)
Group III (*Analyst, Analytical Chemistry, Discussions of the Faraday Society, Journal of Physical Chemistry, Journal of Research of the National Bureau of Standards,* and *Transactions of the Faraday Society*)	4.1% (3.40%, 4.85%)	9.8% (8.64%, 11.02%)	14.4% (12.52%, 16.39%)

The elimination of short titles between 1958 and 1968, however, was not as large as for group I. The difference for group II was observed between 1948 and 1958. Of the three journals in this group, *Journal of the American Chemical Society* rated among the best also in criteria *A* through *E*.

The third group of journals had the lowest mean value of short titles (T_{123}) in 1948, 1958, and 1968, and showed the largest elimination of short titles: from a value of 14.4% in 1948 to a value of 4.1% in 1968. Two of these journals, however, (*Discussions of the Faraday Society* and *Transactions of the Faraday Society*) rated among the lowest for measurements *A* through *E*. This may indicate that elimination of short titles does not necessarily mean that the remaining informative titles have become more informative.

On examining Fig. 2, it might be said with confidence that short titles are definitely being eliminated. It is difficult to say, however, what effect the advent of the KWIC index has had on these changes. The curve for group I is the only one that shows an abrupt decrease in short titles after 1958. Group II shows the opposite effect, as though these journals eliminated their short titles before the KWIC index made its appearance. Group III shows a steady decrease from 1948 to 1968, without any effect that could be attributed to the appearance of permuted-title indexes around 1958.

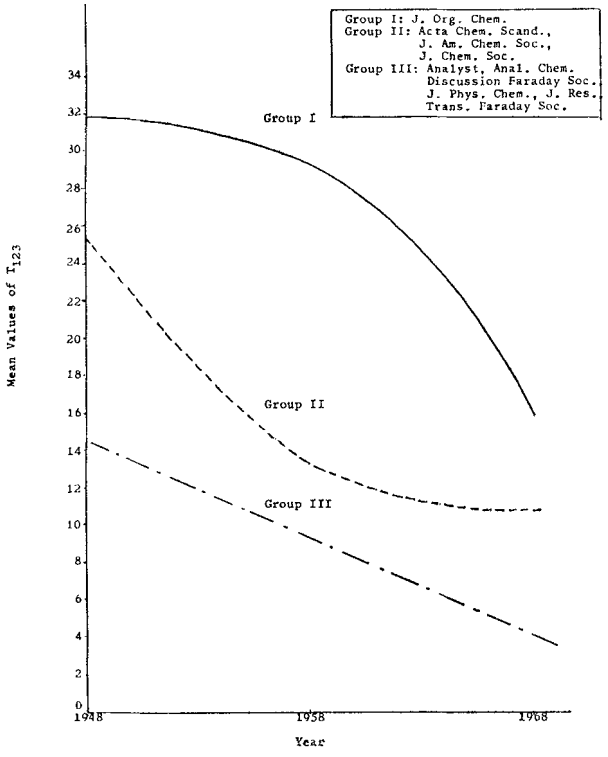


FIG. 2. Mean values of T_{123} (number of titles with three or less substantive words) versus time

Therefore, we can only conclude: *As rated by the criterion adopted, uninformative titles of chemical papers are being eliminated.*

There was a statistically significant decrease in the number of short titles (with three or less substantive words) between 1958 and 1968 and between 1948 and 1958 for all journals tested.

The elimination of short titles in a given journal from 1948 to 1968 does not necessarily correspond to a statistically significant increase in values *A* through *E*. In other words, there are at least two distinct changes that may occur in titles: Informative titles may become more informative, and uninformative titles may be eliminated. These two changes do not necessarily coincide.

• Limitations of This Study

The differences observed when comparing changes in titles of chemical papers suggest that journals fall into certain patterns. The homogeneity of the field covered, the degree of standardization of the particular terminology, the average length of terms particular to the field, and the editorial policy of the journal are but a few of the factors that make the interpretation of the observed patterns difficult. It is not possible, therefore, to draw a conclusion from the differences observed among journals. Some, such as *Journal of Organic Chemistry,* strongly suggest that the field of chemistry is too heterogeneous and should be subdivided in further studies of this kind to be more meaningful.

Underlying our measurements *B* and *C* (matches made between title and abstract) is the following assumption: For a given journal, the abstracts accompanying an article have remained essentially the same in quality from 1948 to 1968. The abstract has been considered a constant entity against which potentially changing titles were compared. It is probable that some changes have occured also in the abstract from 1948 to 1968, but such changes are difficult to measure or estimate.

Decisions concerning whether a word is to be counted as a substantive word, whether two words are to be considered synonyms, or whether two hyphenated words constitute one or two substantive words are inherently subjective. The task of writing explicit and unambiguous rules to be followed in these decisions posed a problem. Stating these rules with clarity and precision for possible replication of the study by another worker was found difficult.

• Conclusion

The statistical significance of the differences we measured is such that we can say with confidence: "Yes, titles of chemical papers are becoming more informative!"