Key Word versus Key Phrase in Manual Indexing

Mohammad Reza Falahati Qadimi Fumani
Regional Information Center for Science and Technology, Shiraz
Iran
mrfalahat@yahoo.com

Abstract: The main purpose of this article was to check variations in the application of each type of key term – 1-word, 2-word, 3-word, 4-word, and phrase – among four groups, three manual indexers, and 111 authors; to analyze the application of single word versus 2-word index terms and single word versus phrase index terms by each of the four groups. The analyses carried out with the SPSS software bore the following results: (1) The results of one-way ANOVA revealed that the four groups of indexers were statistically different in their application of each type of key term; (2) All the four groups of indexers used phrases more than single word index terms; (3) Each of the three indexers and authors introduced 2-word key terms more than 1-word key terms.

I. Introduction

In the International Encyclopedia of Information and Library Science (IEILS, 2003, p. 341) "keyword" was defined as "A word that succinctly and accurately describes the subject, or an aspect of the subject, discussed in a document." In order to select keywords, librarians may use "subject headings", which have the double functions of describing the content of a document, and grouping library headings, that is, all documents related to a subject area are grouped together. This will facilitate users' access to documents related to a given subject.

Besides subject headings, there are other sources from which key terms can be extracted like a "controlled vocabulary", which embodies a standardized set of descriptors. Unlike the two categories introduced above (subject headings and controlled vocabularies), key terms could also be extracted from a third source, say, the "body" of the document itself. For instance, KWIC, KWAC and KWOC are all indexing techniques that work only based on the lexical information available in the title of the document. "Full-text indexing" is also another method where the whole body of the document is stored for retrieval purposes. Full-text indexing is, of course, disadvantageous because it requires huge memory space and therefore is not suitable for large

datasets. Key term indexing can therefore be viewed as advantageous over full-text indexing since the former requires much smaller memory space.

Both single words (keywords) and phrases (key phrases) may be referred to as "key terms", "content tags" or "content labels", which are all aiming at representing the content of some part of an article or document.

A keyword embodies only a single lexical item like "book" and "rice". Zhai et al. (1997, pp. 347-348) stated that single words, as index terms, may have two problems: Firstly, they may be misleading. For example, in lexical atoms like "hot dog", the contained single words do not carry their regular meanings and are therefore very misleading if employed as separate indexing terms. Secondly, they may be too general. For instance, the individual words "junior" and "college" are not specific enough to distinguish "college junior" from "junior college".

A key phrase is a key term that includes at least two lexical items. Hulth (2004) found out that more than 86% of the whole key terms introduced by indexers were phrases or "multi-word lexemes" as she called them. Silverstein et al. (1998) found out that in Altavista the average length of each query was 2.4 terms. Engl et al. (1997) reported that 30% of the terms keyed in as search queries by users were phrases and that users considered key phrases as important in leading them towards their information needs as traditional methods like using author name and title search, e.g., KWIC, KWOC, KWAC. Medelyan (2005, p. 4) believed that shorter phrases could organize small collections of documents into larger thematic groups, whereas longer phrases could describe the document's content more precisely. Assessing the browsing behavior of the Internet users also unveils that when they want to search for some piece of information, they frequently prefer to use longer phrases on belief that the documents retrieved will be more relevant and to the point.

Key phrases have also proved useful in navigation (Gutwin et al. 1998), thesaurus construction (Paynter et al. 2000), text clustering and classification, automatic text summarization, and in many other areas. The effectiveness of key phrases depends, to a great extent, on how well and how carefully they have been assigned. Those assigned by experienced human indexers are, in

general, believed to be the best ones, although this does not mean that all experienced human indexers are the same and produce identical or necessarily qualitative indexes.

II. Objectives of the Study

In this article, the length of each single index term, rather than the length of each key term log, as introduced by the three indexers and authors of the .doc articles (totally four groups), was targeted as the subject of the analysis. The main objective was to check variations in the application of each type of key term – 1-word, 2-word, 3-word, 4-word, and phrase – among four groups, three manual indexers, and 111 authors. Two other objectives were to analyze the application of single word versus 2-word index terms, and single word versus phrase index terms by each of the four groups.

III. Hypotheses

Based on the above objectives of the study, the following hypotheses were formulated:

$H_1.1$: The three human indexers and the authors of 111 .doc articles assign key terms of various lengths differently.

> **$H_1.1.1$:** The three human indexers and the authors of 111 .doc articles are different in their application of each type of key term — 1-word, 2-word, 3-word, 4-word, and phrase.
>
> **$H_1.1.2$:** Each of the three human indexers and the authors of the 111 .doc articles use 1-word versus phrase key terms differently.
>
> **$H_1.1.3$:** Each of the three human indexers and the authors of 111 .doc articles use 1-word versus 2-word key terms differently.

IV. Procedure

To carry out the study, the following steps were taken:

1. The key terms introduced by the authors of the 111 .doc articles (all related to the domain of agriculture) as well as those introduced for the same articles by the three human indexers

(each with an M.A. degree in Librarianship and Information Science and with at least five years of indexing experience) using a controlled vocabulary (Persian agrovoc) were reviewed.

2. Each key term was classified in one of the five categories, namely 1-word, 2-word, 3-word, 4-word, and phrase.

3. Descriptive and inferential statistics including frequency tables as well as paired sample t-test and one-way ANOVA were employed to study variations and differences in the application of key terms of various lengths as observed by each group and among the four groups.

V. Data Analysis

The analysis related to each hypothesis will be provided separately.

$H_1.1.1$: *The three human indexers and the authors of 111 .doc articles are different in their application of each type of key term — 1-word, 2-word, 3-word, 4-word, and phrase.*

This hypothesis was the main objective of the present study and aimed at finding whether the three human indexers and the 111 authors were different in their application of each type of key term — 1-word, 2-word, 3-word, 4-word, and phrase. Here, "phrase", and accordingly "key phrase", was defined as a term consisting of more than one free morpheme. For example, ' جنگل ها ' (TC=3062) [Jngl*hA] /dʒæŋgæl*hɑ:/ "Forests" was considered to be a single word, but ' جنگل های بارانی ' (TC=35654) [Jngl*hA.y+bA.r.A.ny] /dʒæŋgæl*hɑ:jɛ+bɑ:rɑ:ni:/ "Rain forests" was considered as a phrase consisting of two terms. Morphemes joined together to express a single concept were also considered as single terms, e.g., ' کار کردن ' [kA.r+kr.d.n] /kɑ:r+kærdæn/ "To work". Further, terms consisting of one free morpheme and a bound morpheme were considered single word terms, e.g., the two morphemes ' برنامه ' [br.nA.mh] /bærnɑ:mɛ/ and ' ریزی ' [r.yz.y] /ri:zi:/ in ' برنامه ریزی بازگشتی ' (TC=28946) [br.nA.mh*r.y.z.y+bA.z.gshty] /bærnɑ:mɛ*ri:zi:jɛ+bɑ:zgæʃti/ "Recursive programming" were considered as a single word.

Later, this surface form made a 2-word phrase by attaching to 'بازگشتی' [bA.z.g<u>sh</u>ty] /bɑ:zgæʃti/ "Recursive".

Table 1: Frequency and Percentage of Key Terms of Various Lengths Introduced by the Four Groups.

Length of Key Term	Indexer 1		Indexer 2		Indexer 3		Authors.doc	
	Sum	**%**	**Sum**	**%**	**Sum**	**%**	**Sum**	**%**
1-word	400	31	309	41	314	32	171	37
2-word	731	57	378	50	526	54	222	48
3-word	123	10	59	8	125	13	56	12
4-word	25	2	13	1	6	1	16	3
Phrase*	879	69	450	59	657	68	294	63
Total	1279	100	759	100	971	100	465	100

Note: "Phrase" here refers to all groups except the 1-word class; Valid N for each of the four groups discussed is 111.

Table 1 has summarized the frequency and percentage of each type of key term as produced by the three human indexers as well as authors of the .doc articles.

In Table 1, 2-word phrases were most common in the key terms introduced by the four groups. In all, 731 (57%) of the key terms assigned by indexer 1 and 378 (50%) and 526 (54%) of those assigned by indexers 2 and 3 were 2-word phrases. For the authors of .doc articles, this rate was 222 (48%). In all the four groups, single word index terms ranked second with 400 (31%), 309 (41%), 314 (32%), and 171 (37%) for indexers 1-3 and authors respectively, and 4-word phrases were the least frequent index terms. Further, single word and 2-word key terms together comprised the main part of the index terms introduced by each group. The rates computed were 1,131 (88%), 687 (91%), 840 (86%) and 393 (85%) for indexers 1 to 3 and authors. In all, the four groups introduced 69%, 59%, 68% and 63% phrases respectively.

Table 2: Frequency and Percentage of Key Terms of Various Lengths Introduced by the Three Human Indexers as a Single Group.

Length of Key Term	Indexers 1-3	
	Sum	**%**
1-word	1023	34%
2-word	1635	54%
3-word	307	10%
4-word	44	2%
Phrase*	1986	66%*
Total	3009	100%
Valid N	333	

Note: The "phrase" class denotes 2-word, 3-word and 4-word index terms. Further, the information related to the "phrase" class has not been included in the numbers reported in the "total" row.

In Table 2, the data related to the three human indexers were considered together as a single group. The results indicated that 2-word phrases were employed more than half (54%) of all key terms assigned by the three indexers, and single word, 3-word and 4-word key terms ranked second to fourth with 1,023 (34%), 307 (10%), and 44 (2%) respectively. In all, 1986 (66%) of the index terms they assigned were phrases.

Table 3: Mean and Standard Deviation of Key Terms of Various Lengths as Produced by the Three Indexers and Authors of .doc Articles.

Indexers		1-word	2-word	3-word	4-word	Phrase
1	**Mean**	3.60	6.58	1.10	.22	7.9
	SD	2.08143	2.84275	1.16269	.41963	3.31973
2	**Mean**	2.78	3.40	.53	.11	4.04
	SD	1.45481	1.69157	.78416	.32302	1.75208
3	**Mean**	2.82	4.73	1.12	.05	5.9
	SD	1.66719	2.10934	1.16866	.22715	2.28487
Authors .doc	**Mean**	1.54	2.00	.50	.14	2.64
	SD	1.36037	1.24316	.73699	.37772	1.25516
Total	**Mean**	2.68	4.18	.81	.13	5.12
	SD	1.81610	2.65989	1.02588	.34879	3.02058

Note: Valid N for the total group (last row) is 444 and for each of the other four groups is 111.

Table 3 has summarized the mean and standard deviation of each type of key term for the three human indexers and the 111 authors of .doc articles.

In order to find out whether the mean differences observed above between the four groups as regards the application of each type of key term were statistically significant, one-way ANOVA was employed (Table 4).

Table 4: Results of One-way ANOVA to Test the Significance of Mean Differences as Regards the Application of Each Key Term Type by the Four Groups.

Length of Key Term	Groups	Sum of Squares	df	Mean Square	F	Sig.
1-word	**Between Groups**	242.423	3	80.808	29.175	.000 HS
	Within Groups	1218.685	440	2.770		
	Total	1461.108	443			
2-word	**Between Groups**	1271.106	3	423.702	100.063	.000 HS
	Within Groups	1863.117	440	4.234		
	Total	3134.223	443			
3-word	**Between Groups**	39.899	3	13.300	13.726	.000 HS
	Within Groups	426.324	440	.969		
	Total	466.223	443			
4-word	**Between Groups**	1.676	3	.559	4.707	.003 S
	Within Groups	52.216	440	.119		
	Total	53.892	443			
Phrase	**Between Groups**	1744.378	3	581.459	111.356	.000 HS
	Within Groups	2297.514	440	5.222		
	Total	4041.892	443			

The results of one-way ANOVA (Table 4) revealed that the four groups of indexers were statistically different in their application of each type of key term. The differences observed among the four groups were significant for 4-word key terms ($F = 4.707$, $p = .003$), and highly significant for single word key terms ($F = 29.175$, $p = .000$), 2-word ($F = 100.063$, $p = .000$), 3-word ($F = 13.726$, $p = .000$) and phrases ($F = 111.356$, $p = .000$).

Based on the findings, the hypothesis "The three human indexers and the authors of the 111 .doc articles assign key terms of various lengths differently." was accepted for all the five types of key terms.

Table 5: Results of Scheffe Test to Trace Mean Differences in the Application of Key Terms of Various Lengths by the Four Groups of Indexers.

Length of Key Term	Indexers	Indexer1	Indexer2	Indexer3	Authors .doc
1-word	**1**	--	.004 S	.008 S	.000 HS
	2		--	.998 NS	.000 HS
	3			--	.888 NS
	Authors.doc				--
2-word	**1**	--	.000 HS	.000 HS	.000 HS
	2		--	.000 HS	.000 HS
	3			--	.000 HS
	Authors.doc				--
3-word	**1**	--	.000 HS	.999 NS	.000 HS
	2		--	.000 HS	.998 NS
	3			--	.000 HS
	Authors.doc				--
4-word	**1**	--	.142 NS	.004 S	.381 NS
	2		--	.602 NS	.952 NS
	3			--	.286 NS
	Authors.doc				--
Phrase	**1**	--	.000 HS	.000 HS	.000 HS
	2		--	.000 HS	.000 HS
	3			--	.000 HS
	Authors.doc				--

*. Uses harmonic mean sample size= 111.

In order to trace the significance of mean differences, Scheffe test was employed as post hoc of the one-way ANOVA. Table 5 has summarized, in the form of matrix, the results of all the Scheffe tests computed. For each of the five types of index terms, six comparisons were made. The pairs compared included: Indexers 1 and 2, 1 and 3, 2 and 3, 1 and authors, 2 and authors, and finally 3 and authors.

The results indicated that all the four groups of indexers revealed a highly significant difference in their application of 2-word ($p = .000$) as well as phrase index terms ($p = .000$). For single word index terms, non-significant differences were observed only in two pairs, between indexers 2 and 3 ($p = .998$), and indexer 3 and authors ($p = .888$). For the rest of the pairs, a significant difference was observed at the .01 level.

For 3-word index terms, a highly significant difference was observed in four pairs ($p = .000$).
The differences were not, however, statistically significant in two pairs, namely indexers 1 and 3
($p = .999$) and indexer 2 and authors ($p = .998$).

For 4-word index terms, only in one comparison, the one between indexers 1 and 3, a significant
difference ($p = .004$) was observed. For the other 5 comparisons, non-significant differences
were found, meaning that the duo in each pair performed similarly in their application of 4-word
index terms.

The overall conclusion was that quite often indexers differed in their application of each type of
key term. In 21 out of 30 comparisons made in Table 5, a significant difference was observed
between the duo in each pair.

*$H_1.1.2$: Each of the three human indexers and the authors of the 111 .doc articles use 1-word
versus phrase key terms differently.*

The second hypothesis intended to check the application of 1-word versus phrase key terms (all
types of key terms together except the 1-word terms) by each of the four groups. A paired sample
t-test was computed for each of the four groups as illustrated in Table 6.

Table 6: Paired Sample t-tests for the Application of 1-word versus Phrase Key Terms by the
Three Indexers and Authors.

Groups		Mean	SD	t	df	Sig.
Indexer 1	**1-Word vs. Phrase**	4.31	4.20818	10.804	110	.000 HS
Indexer 2		1.2703	2.91687	4.588	110	.000 HS
Indexer 3		3.0901	3.29338	9.885	110	.000 HS
Authors		1.1081	2.30553	5.064	110	.000 HS

Note: Phrase = 2-word + 3-word + 4-word index terms.

Table 6 indicated that indexer 1 ($t = 10.804$, $p = .000$), indexer 2 ($t = 4.588$, $p = .000$), indexer 3
($t = 9.885$, $p = .000$) and authors ($t = 5.064$, $p = .000$) introduced phrases more than 1-word index
terms. The differences observed were highly significant in all cases. An overview of the mean
scores in Table 3 indicated that in each group the mean score obtained for phrase was higher than

that obtained for 1-word index terms: (7.9 versus 3.6 for indexer 1; 4.04 versus 2.78 for indexer 2; 5.9 versus 2.82 for indexer 3, and 2.64 versus 1.54 for authors).

Based on the findings, the hypothesis "The three human indexers and the authors of the 111 .doc articles are different in their application of each type of key term — 1-word, 2-word, 3-word, 4-word, and phrase." was accepted, implying the higher application of phrase key terms.

H1.1.3: Each of the three human indexers and the authors of 111 .doc articles use 1-word versus 2-word key terms differently.

The main intention here was to compare the application of 1-word versus 2-word key terms by each group (Table 7).

Table 7: Paired Sample T-tests for the Application of 1-word versus 2-word Key Terms by the Four Groups.

Groups		Mean	SD	t	df	Sig.
Indexer 1	**1-word vs. 2-word**	2.9820	3.80904	8.248	110	.000 HS
Indexer 2		.6216	2.78977	2.348	110	.021 S
Indexer 3		1.9099	3.11171	6.467	110	.000 HS
Authors		.4595	2.20241	2.198	110	.030 S

In Table 7, a significant difference was observed in the application of 1-word versus 2-word key terms by indexer 2 ($t = 2.348$, $p = .021$), and authors ($t = 2.198$, $p = .03$). For each of the other indexers — indexer 1 ($t = 8.248$, $p = .000$) and indexer 3 ($t = 6.467$, $p = .000$) — the difference observed was highly significant.

Also in Table 3, the mean scores computed for the application of 1-word versus 2-word key terms by each group were 3.6 versus 6.58 for indexer 1; 2.78 versus 3.4 for indexer 2; 2.82 versus 4.73 for indexer 3, and 1.54 versus 2 for authors reiterating that each of the three indexers and authors had introduced 2-word key terms more than 1-word key terms.

Based on the findings, the hypothesis "Each of the three human indexers and the authors of the 111 .doc articles use 1-word versus phrase key terms differently." was accepted, implying the higher application of 2-word key terms.

VI. Conclusion

Human indexers and authors differ in their application of key terms of various lengths. Further, the rate of phrase key terms they assign is significantly higher than that of 1-word key terms. They also use 2-word key terms more than 1-word key terms. The general conclusion to be drawn is that since indexers assign more phrases than words, language processing tools like automatic indexing systems must also assign more phrases. Manual indexers use more phrases as they think phrases carry more meaning compared to single words.

Appendix: Key to Transliteration and Pronunciation

ا [A] /ɑ:/, as in car /cɑ:r/; ب [b] /b/, as in bag /bæg/; ت [t] /t/, as in ten /tɛn/; ج [J] /dʒ/, as in Jack /dʒæk/; د [d] /d/, as in day /dɛi/; ر [r] /r/, as in red /rɛd/; ز [z] /z/, as in zoo /zu:/; ش [sh] /ʃ/, as in she /ʃi:/; ک [k] /k/, as in camp /kæmp/; گ [g] /g/, as in geese /gi:s/; ل [l] /l/, as in lamp /læmp/; م [m] /m/, as in moon /mu:m/; ن [n] /n/, as in noon /nu:n/; ه [h] /h/, as in hat /hæt/; ی [y] /j/, as in yes /jɛs/; ق [Q] /ɣ/, as in French word merci /mɛɣsi/; /i/, as in see /si:/; /ɛ/ as in ten; /æ/ as in cat; * shows zero space; + shows full space; . shows half space; [] shows transliteration; // shows pronunciation; ' ' encloses the Persian terms; " " encloses the English terms.

References

Engl, D., Friedl, J., Labner, J., Sandnerand, M., Schlacher, W., Schmidt, A., et al. (1997). Schlagwort "benutzerforschung". beobachtungen bei der sachlichen suche im opac des sterreichischen wissenschaftlichen bibliothekenverbundes. *Mitteilungen der Vereinigung sterreichischer Bibliothekarinnen & Bibliothekare.*

Gutwin, C., Paynter, G., Witten, I., NevillManning, C., & Frank, E. (1999). Improving browsing in digital libraries with keyphrase indexes. *Decision Support Systems, 27*(1/2).

Hulth, A. (2004). *Combining machine learning and natural language processing for automatic keyword extraction*. (Doctoral dissertation, Stockholm University, 2004).

Feather, J., & Sturges, P. (Eds.). (2003). *International Encyclopedia of Information and Library Science* (2nd ed.). London; New York: Routledge.

Medelyan, O. (2005). *Automatic keyphrase indexing with a domain-specific thesaurus*. (Master's thesis, Albert-Ludwigs University, 2005).

Paynter, G., Cunningham, S. J., & Witten, I. H. (2000). Evaluating extracted phrases and extending thesauri. In *Proceedings of the Third International Conference on Asian Digital Libraries*, pp. 131-138.

Silverstein, C., Henzinger, H., Marais, H., & Moricz, M. (1998). Analysis of a very large altavista query log. Technical Report 1198-014, Digital System Research Center.

Zhai, C., Xiang, T., Natasa, M., & Evans, D. A. (1997). Evaluation of syntactic phrase indexing — CLARIT NLP Track Report. In E. M. Voorhees & D. K. Harman (Eds.). *The 5th Text Retrieval Conference (TREC-5)* (pp. 347-358). NIST Special Publication 500-238.

Author:
Mohammad Reza Falahati Qadimi Fumani, Ph.D. in Computational Linguistics, /Mohæmmæd rɛzɑ: fælɑ:hæti: ɤædi:mi: fu:mæni:/, Faculty of the Computational Linguistics Research Department, RICeST, Shiraz, Iran