Keyword Analysis of Intellectual Capital and Knowledge Management in SCOPUS

Aitouche Samia, Laggoune Assia, Titah Mawloud, Mouss Mohamed Djamel, Zerari Naima, Latreche Khaled, Kanit Akila and kanit Abdelghafour, Department of Industrial Engineering, Laboratory of Automation and Manufacturing, University Batna 2, Batna, Algeria

email: samiaaitouche@yahoo.fr, a.laggoun12@yahoo.fr, t.mawloud@yahoo.fr, d_mouss@yahoo.fr, zerari99@yahoo.fr, khaled_lat@yahoo.fr akila.msi2016@gmail.com, k_abdelghafour@yahoo.fr

Abstract— **The aim of this paper is to perform a keyword analysis in two areas of research: Intellectual Capital and Knowledge Management. The keywords are of three types : keywords proposed by the authors in their articles, the keywords that users use in their queries and the densest words existing in a textual corpus. Zipf's law usually applied for natural language is applied in this work for scientifc corpus constituted of confused full articles in each area. We wrote 8 R programs going through titles of articles, authors' keywords, abstracts and full articles to calculate frequencies and interpret them. The keywords of intellectual capital measurement and diclosure have the highest frequencies. The measures are stated by companies in annual reports and could not be integrated in their balance sheets because the classical accounting does not take into acount intellectual capital as an asset. Knowledge management is more oriented towards the capitalization of knowledge to improve business performance and for industries. This work is the first in keyword analysis for the two areas. It could be usefull to prepare glossaries, ontologies and all semantic reasearhes of research areas.**

Keywords-content analysis; keyword analysis; dense word; Zipf's law; intellectual capital; knowledge management.

I. INTRODUCTION

Our work is a quantitative analysis of keywords and in the areas of intellectual capital (IC) and knowledge management (KM). IC focuses on building and governing intellectual assets from strategic and enterprise governance perspectives with some focus on tactics [1]. KM is more detailed and focuses on facilitating and managing knowledge related activities such as creation, capture, transformation and use [1]. Otherwise, they are two faces of the same thing (knowledge); IC capitalizes it to create value and wealth and, KM manages it from acquisition to diffusion.

A set of keywords indexes and densest words in a scientific article are proposed, calculated and analyzed to have an idea on the trends of scientific area via most used and extracted keywords and their evolution in time. In this paper, Zipf's law was applied to confused abstracts of articles of IC and KM. The results were not satisfactory because the size of abstracts' corpus is not sufficient to apply Zipf's law, which is more adapted to larger corpuses. So we applied it on confused full articles of IC and KM separately to try to reach better results.

The gap that we tried to satisfy is the inexistence of keyword analysis in IC and KM. This analysis is limited to articles existing in SCOPUS database and could be extended to other databases in the future.

The paper is structured as follows: Section II contains the related works and Section III designs the solution and the application of R programs for the keyword analysis of IC and KM areas. Section IV is dedicated to analysis of results and finally, possible extensions to the work are announced to conclude the paper in Section V.

II. STATE OF THE ART

In literature, there are several goals to keyword analysis. Among them, there is the research of Jaime I. L. et al [2] which is a proposition of a new keyword search algorithm that takes into account the semantic information extracted from the schemes of the structured and semi-structured data sources. Bang [3] focuses on the disciplines of the journal information system frontiers (ISF) researches. The author created a keyword classification scheme, incorporating new research topics into Barki's information systems keyword classification scheme [24], to describe the disciplines of ISF until 2012, examining word frequency and keyword co-occurrence. This work is limited to one journal in information systems and its results are not generalizable. On the other hand, Choi and Kang [4] extracted manually keywords from abstracts and titles from Journal of Educational Technology (JET) between 1985 and 2013 and found that educational technology research in Korea has been strongly influenced by new media, design theory, and educational assessment. The manual extraction of keywords takes time, limits the sample of articles and statements are true only for the sample; one journal is not sufficient. Wu Bihu et al [5] examine author-selected keywords of research published in Annals of Tourism Research. In total, 5534 keywords from 2504 articles form the basis of this analysis. Iterative coding results in 200 core keywords serving as descriptors of major research subjects, and 10 gene words indicating knowledge domains formed through cross-references and hybridization of core keywords. By employing the social network analysis technique, Gohar and Jacob [6] found that the results highlight the importance of digital media and business, and IT governance to today's information technology management environment.

Zipf's law is used in this work to show how much a scientific text respects it. Zipf stated that if one takes the words making up an extended body of text and ranks them by frequency of occurrence, then the rank of words multiplied by their frequency of occurrence will be approximately constant [7]. In [8], it is shown that the distribution of word frequencies for randomly generated texts is very similar to Zipf's law observed in natural languages such as English. On the internet, Zipf's law appears to be the rule rather than an exception. It is present at the level of routers transmitting data from one geographic location to another and in the content of the World Wide Web [9].

Our work provides a keyword analysis based on the calculation of frequencies of keywords in titles, abstracts and content of scientific articles for IC and KM areas to analyze quantitatively their trends in terms of themes, expressed by these frequencies.

III. METHODOLOGY AND DATA

Databases of research as SCOPUS do not offer keyword analysis and impose the type of exported data files; the performed analyses were adapted accordingly.

A. *Analysed data*

The data is exported from SCOPUS, i.e., a database which indexes thousands of scientific journals and more than fifty millions scientific articles in all areas of research. The data is a set of articles with "intellectual capital" or "knowledge management" in their titles. Analyses were performed on abstracts and full papers (PDF files). The scientific production is higher in KM (10000 articles) than IC (1500 articles), according to data exported on April 16th 2016.

B. *Methodology*

Eight different programs are written in R language to achieve objectives of the study. The algorithms of the frequencies of author's keywords and Zipf's law are presented as examples; they produce results for sections A and E respectively after their implementation. The six remaining programs are developed for the results of B, C and D sections.

Algorithm frequency_authors_keywords

The authors' keywords algorithm contains generic steps

Begin

1- Read the Excel file entries,
2- Separate authors' keywords of that article and put it in a table as they are combined in the same Excel box and separated by semicolons
3- Repeat 1-2 until the Excel file is ended
4- Calculate the frequency of authors' keywords
5- Order by descending order of frequencies
6- Export the ordered Table in Excel
7- Draw the overall graph on years

End.

Algorithm Zipfs_law

The full articles are downloaded and gathered in the same directory to make a text corpus of an area.

Begin

1- Create an Excel file1 containing the names of PDF (acrobat reader) files gathered in a directory, each name in a row,
2- Read the first name of PDF file from the Excel file1 and use it to open the PDF file,
3- Read word by word and write them in text file3 created to include all the content of all PDF files together, and at the end of this operation it will be constituted the corpus of all PDF files (text corpus),
4- Omit the punctuation from the text file3
5- Repeat instructions 2-3 until the Excel file1 is ended,
6- Each word of corpus from text file3 is added in a row in a second Excel file2 without repetition of words,
7 - For each word from Excel file2, calculate frequency of the word in the corpus of text file3
8- Repeat the operations 6-7 until text file3 is ended,
9- Order the Excel file2 in descending order of frequencies
10- For each word from Excel file2, calculate the value of the existing Zipf's law for this word and put it in the second column, using the formula:
Existing Zipf's law (word) = Frequency (word) * rank (word in file2),
11- Calculate the theoretic value of Zipf's law for each word and put it in the third column, only the size of corpus is used as input to the theoretic value using the formula:
Theoretic Zipf's law (word) = Size of the corpus of text file3/rank (word in Excel file2),
12- Repeat 10-11 until the end of Excel file3,
13- Draw the graphs of theoretic and existing Zipf's law
- Note: this algorithm is applied once on IC and once on KM corpuses separately.

End.

IV. RESULTS AND DISCUSSION

The analyses were performed on abstracts, full papers and the other items of paper as titles and authors' keywords. In certain cases, papers were analyzed individually then collectively.

A. *Authors' keywords analysis*

1) Authors' keywords analysis in IC area: In TABLE I, we present the obtained frequencies of the authors' keywords, in descending order (**2308** keywords). The keywords highlited in TABLE I are discussed.

TABLE I: FREQUENCIES OF AUTHORS' KEYWORDS IN IC AREA.

Authors' keywords	freq
Intellectual capital management	**865**
Intellectual capital disclosure	860
Intellectual capital statements	850
Intellectual capital reporting	850
Intellectual capital	847
Intellectual capital assets	847
Intellectual capital development	845
Intellectual capital efficiency	839
Intellectual capital redefinition	835
Intellectual capital dimensions	820
Intellectualcapital's components	812
Intellectual capital evaluation	808
Intellectual capital performance	**808**
Intellectual capital of organizations	764
Measuring, management and Reporting Intellectual	**764**
Intellectual capital measurement	**758**

It is noted that research in IC area is oriented more towards **IC management**. At first, accountants are conscious about the measurement of IC as intangible asset and tried to incorporate it to classical balance sheet next to tangible assets, but they have failed. Nowadays, the interest is extended to **managers** trying to disclose and exploit IC to improve business **performance** and IC is reported separately. In addition, it is noted that authors of IC area use a relatively short list of keywords to express the content of their scientific work but with significant frequencies and slight degradation between them. This is explained by the novelty of the field.

2) Authors' keywords analysis in KM area: for KM, the frequencies of authors' keywords are represented in TABLE II, in descending order (**6650** keywords).

TABLE II. FREQUENCIES OF AUTHORS' KEYWORDS IN KM AREA

Authors' keywords	freq	Authors' keywords	freq
knowledge management	**271**	Tacit knowledge	222
Competitive advantage	**253**	information retieval	217
data mining	**252**	knowledge management (KM)	215
knowledge sharing	**252**	Mathematical models	215
humain ressource	**237**	Technology	215
artificial intelligence	236	organisational performance	120
Article	226	**Manufacture**	**119**
world wide web	226	health care	118
strategic planning	222	….	….

It is noted that researches in KM are oriented more toward **KM systems.** The possession of tacit knowledge by experts is a **competitive advantage**. **Data mining** is used for automatic extraction of knowledge from existing databases. **Knowledge sharing** improves collaboration and is a concern of **human resources management** in a business or in a particular production system. The authors' keywords **Manufacture, industrial engineering, industrial economics and cybernetics** are a strong argument that KM systems support industrial engineering and general industry. Moreover, it is noted that the authors of KM use a wide list of keywords relatively to express the content of their scientific work but with less frequencies than those of IC and have rapid degradations in frequencies.

B. *Analysis of users' keywords in the summaries and complete articles*

The purpose of this analysis is to go deeper in the body of articles and search the frequencies of keywords requested by a user (could be scientist) in abstracts or full articles, to choose articles more appropriated to these keywords and incorporate them in his work (e.g. in the form of literature review).

1) Analysis of users' keywords in abstracts of IC area: An individual analysis of articles' abstracts is performed. It is beneficial when the user wishes select an abstract or a group of abstracts to read. The frequencies of users' keywords when analyzing abstracts one by one, vary from an abstract to another and from keyword to another. For reasons of clarity and space, we presented a small sample most frequent keyword requested. For better analysis, cumulative frequencies of confused abstracts are presented in TABLE III.

TABLE III. CUMULATIVE FREQUENCIES OF USER KEYWORDS IN ABSTRACTS OF IC

User keyw	Freq	User keyw	Fre	User keyw	Freq
Value	**1768**	**qualitative**	**144**	Sheet	46
industry	**632**	Risk	140	Score	42
market	**567**	**industrial**	**126**	Danish	36
empirical	556	**quantitative**	**115**	skandia	22
measurement	550	guidelines	114	cement	4
method	462	Wealth	104	**VAIC**	**0**
accounting	356	scorecard	70	**NICI**	**0**
Added	254	balanced	54	**IC-dVal**	**0**
qualitative	**144**	**…**	**…**	**WWTK**	**0**

The most frequent words are **market** and **value**. It is explained by the relationship between IC and the market value in certain methods of IC measurement (IC value = Market value - book value). It represents generally 4 or 5 times the book value (value of business given by accountants), which explains the existence of a very important intrinsic value that is *intellectual capital* in the form of the know-how of the experts in business. The words **industry** and **industrial** are very frequent; this implies that IC is measured also within industrial businesses. The **qualitative** word is better ranked than **quantitative** one because of the qualitative nature of IC which is difficult to quantify. The managers and accountants look for a credible tool to measure it. **VAIC**, **NICI**, **IC-dVal** and **WWTK** are user requested measurement methods of IC, but do not exist in the sample of articles containing "intellectual capital" in their titles.

2) Analysis of users' keywords in abstracts of KM area: The results of TABLE IV shows the cumulative frequencies of keywords in confused abstracts of KM.

TABLE IV: THE CUMULATIVE FREQUENCIES OF USERS' KEYWORDS IN KM

User keyw	Freq	User keyw	Freq	User keyw	Freq
Data	**782**	**potential**	**116**	Logic	22
performance	**772**	**intelligence**	**100**	Neural	12
sharing	**432**	**industrial**	**97**	Genetic	7
industry	**250**	Risk	93	Cement	**2**
practice	**229**	**communities**	**71**	commonKads	1
manufacturing	**171**	Fuzzy	67	datawarehouse	1
ontology	**157**	Database	60	**MSKM**	**0**
community	**136**	Center	35	**MASK**	**0**
Network	**133**	**…**	**…**	**…**	**…**

The word **data** is the densest, it expresses the close relationship between knowledge and data [15]; the data when it is processed becomes information and when it is used it becomes knowledge. The word **performance** reflects that knowledge contributes to improve performance of a company [16]. The word **sharing** expresses the sharing of knowledge, reinforces the collaboration between members of professional team and gives better results than the individual tasks. This collaboration is carried out by communities of practice (**community**, **communities**, **practice**); they are groups of employees using knowledge in their work. The industry is also present in the KM (**industry**, **manufacturing and industrial**). The word **intelligence** argues the relationship between **KM** and business intelligence. It is involved in enriching the knowledge base. **MSKM** and **MASK** are two methods of knowledge management but don't exist in the studied sample containing only articles with "knowledge management" in their titles.

3) Analysis of users' keywords in the full articles in KM: We used the PDF files of articles downloaded from SCOPUS. Fig. 1 shows the results when analyzing file by file. The red surrounded area represents the highest frequencies of users' keywords. It could be useful to a researcher looking for a specific topic by selecting only the articles with high frequencies of required keyword (topic) to use them in their literature review.

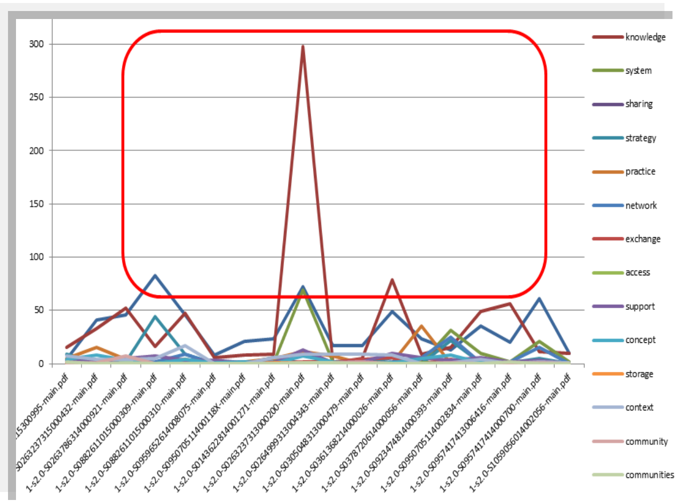

Figure 1. Frequencies of user keywords in PDF files individually in KM for 2015

In this individual analysis by PDF file, frequencies of keywords differ from one file to another. A global analysis would be more significant (TABLE V) when the researcher looks for the best area treating its topic (keyword) or to compare topics in the same area (IC, KM…etc).

TABLE V: GLOBAL USERS' KEYWORDS MEANS OF FREQUENCIES IN PDF KM IN 2015

User keyw	Freq	User keyw	Freq	User keyw	Freq
management	32,72	**Context**	**4,78**	Sharing	0,94
knowledge	39,94	**Network**	**3,94**	community	0,78
System	**8,5**	Support	3,94	Access	0,5
strategy	**5,94**	Concept	2,67	communities	0,39
practice	5,11	exchange	1,22	Storage	0,11

The word **system** is the most frequent; it relates to knowledge management systems. The word **context** refers to the contextual nature of knowledge and its extraction by data mining. **Network** refers to its important role in communication and collaboration to share knowledge by the communities of practice.

C. Analysis of the densest words

It is called density of words in the abstracts, the frequency of any word in a text corpus, not keywords provided by a user or by author. This analysis complements the previous one, because sometimes there are important new words discovered by this analysis but less known by the user. It is an automatic extraction of the densest words. Excel files containing only abstracts are used. In other words, it consists to find frequencies of words in a corpus filtered from stopwords which are in English (a, an, the, then, it, he, she, about, etc.), commonly conjunctions, pronouns or any other word not affecting the meaning of a text. The set of stopwords exists in the literature and it is taken into account in certain languages as R; the used language in this work.

1) Analysis of the densest words in the abstracts

a) Analysis of the densest words in the abstracts of IC: The means of the densest words in the IC area are calculated from 1995 to 2016. In all years, the word **intellectual** and **capital** have approximately the same appearance, the frequency of **capital** is more than **intellectual** (TABLE VI) because it exists other capitals: human capital, structural capital and organizational capital…etc.

TABLE VI: FREQUENCIES' MEANS OF DENSEST WORDS IN ABSTRACTS IC.

Dense word	Mean	Dense word	Mean	Dense word	Mean
Capital	**8549**	**managemen**	**1920**	findings	1057
intellectual	**6441**	**Value**	**1530**	information	906
Paper	**2181**	**companies**	**1282**	Can	893
Study	**2176**	**purpose**	**1199**	Limited	499
knowledge	**2067**	Firms	1120	creation	420
Research	**2007**	Model	1082	**indicators**	**376**
performance	**1933**	analysis	1065	implication	55

Paper, **study**, **research**, **findings** and **purpose** are generic words used in abstracts of any paper. **Indicators** express the different **value**s of components of IC used by the measurement **model**s.

b) Analysis of the densest words in the abstracts of KM: TABLE VII shows the averages of frequencies of densest words in KM area from 1997 to 2016. The generic words are same that IC ones. The word **process** indicates the knowledge management process optimization, which is subject to the reactivity of the knowledge system.

TABLE VII. MEAN OF THE DENSEST WORDS IN THE ABSTRACTS OF KM

Dense Word	Mean	Dense Word	Average	Dense Word	Mean
knowledge	1768,83	system	286,33	**innovation**	**156,50**
management	1053,42	paper	219,92	data	123,58
organisational	**268,83**	research	218,58	analysis	102,83
process	334,33	study	188,92	creation	86,33
information	300,33	devlopment	187,58	model	46,08

The word **organisational** is in third rank, it demonstrates the organizational aspect of knowledge management and the encouragement of organizational learning modules included in knowledge management systems. The word **innovation** is present as enrichment element of the knowledge system and the trigger of organisational learning.

2) Analysis of the densest words in articles' titles: The purpose of this analysis is very important, it is a decisive factor on the interest of a scientific article and its content.

a) Analysis of the densest words in the titles of the articles of IC: Fig. 2 shows the mean frequencies of the densest words in the titles of IC from 1998 to 2016.

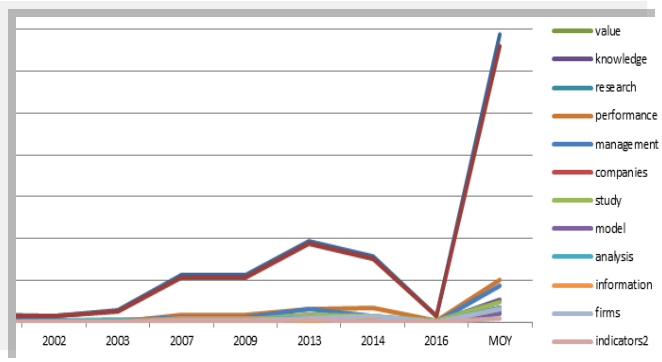

Figure 2. Densest words in the titles of IC, per year

It is noted that the two curves of the words **capital** and **intellectual** look the same on all the years since data is exported from SCOPUS have in their titles **intellectual capital**. The **capital** curve is slightly above the **intellectual** curve because there are other capitals. Their densities are highest in 2012 when the scientific production was highest for IC.

TABLE VIII. MEANS OF DENSEST WORDS IN IC TITLES

Dense word	Mean	Dense word	Mean	Dense word	Mean
capital	687	study	48	information	31
intellectual	659	value	35	model	22
performance	**101**	**analysis**	**34**	**research**	**19**
management	86	**companies**	**34**	**indicators**	**8**
knowledge	53	**firms**	**32**	**…**	**…**

The word **performance** is very dense (Table VIII), because IC improves business performance; it is expressed by the words **companies** and **firms**. This performance is measured by performance **indicator**.

b) Analysis of the densest words in the titles of KM articles: Fig. 3 shows the means frequencies of the dense words in titles of KM from 1997 to 2016.

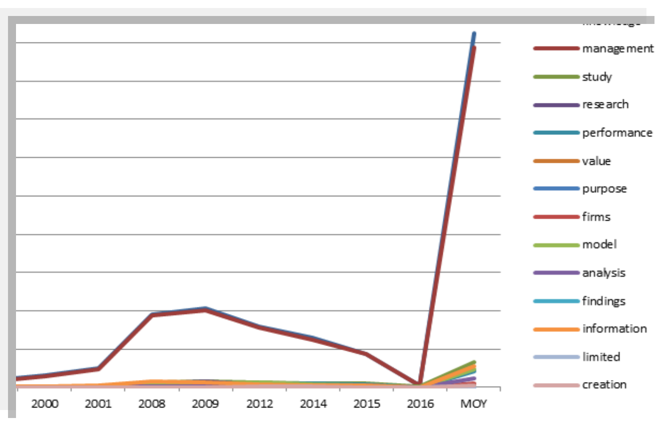

Figure 3. Densest words in titles of KM, per year

The two curves of **knowledge** and **management** look the same on all the years since data is exported from SCOPUS have in their titles knowledge management. The **knowledge** curve is slightly above the curve **management**, contrary to what was found in abstracts and PDF. There are other treatments of knowledge (knowledge acquisition, knowledge sharing…etc.). Their densities are highest in 2009 when scientific production was the highest for KM.

TABLE IX. MEANS OF THE DENSEST WORDS IN KM TITLES

Dense word	Mean	Dense word	Mean	Dense word	Mean
knowledge	4632	research	227	**creation**	**22**
management	4436	**performance**	**214**	findings	8
study	326	analysis	112	purpose	3
information	278	**firms**	**52**	limited	2
model	**240**	**value**	**36**	**…**	**…**

The word **model** expresses the models used for knowledge engineering, knowledge sharing, to improve **performance** of **firms** and assessment of **value** (Table IX).

3) Analysis of the densest words in the authors' keywords: The first analysis performed in Section A, concerns the author keyword as a whole expression, but this analysis seeks the densest words composing this keyword.

a) Analysis of the densest words in the authors' keywords of IC: Fig. 4 shows the frequencies' means of the densest words in the authors' keywords of IC from 1995 to 2016.

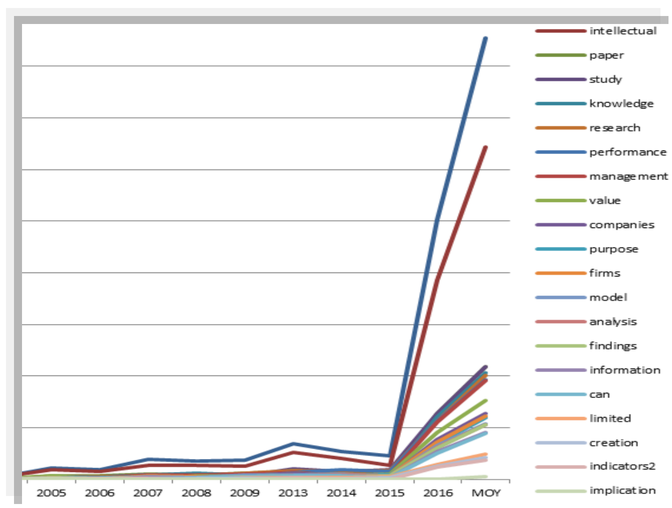

Figure 4. Densest words in authors' keywords of IC, per year

Over the years, the **intellectual** and **capital** curves have almost the same trends over time, but the **capital** one exceeds in density because there are other capitals than **intellectual,** such as financial capital and the components of intellectual capital.

TABLE X THE DENSEST WORDS IN THE AUTHORS' KEYWORDS OF IC

Dense word	Mean	Dense word	Mean	Dense word	Mean
Capital	450	management	101	findings	56
intellectual	339	**Value**	**81**	information	48
Paper	115	companies	67	Can	47
Study	115	purpose	63	Limited	26
knowledge	109	Firms	64	**creation**	**22**
research	106	**Model**	**57**	**Indicators**	**20**
performance	**102**	analysis	56	**implication**	**3**

The results of the density in the authors' keywords are similar to the density in titles, because in general the authors get their keywords from the title. Common dense words (Table X) are **performance**, **value**, **model**, **creation**, **indicator** and **involvement**.

b) Analysis of the densest words in the authors' keywords of KM: Fig. 5 shows the frequencies' means of the densest words in the authors' keywords of KM from 1998 to 2016**.**

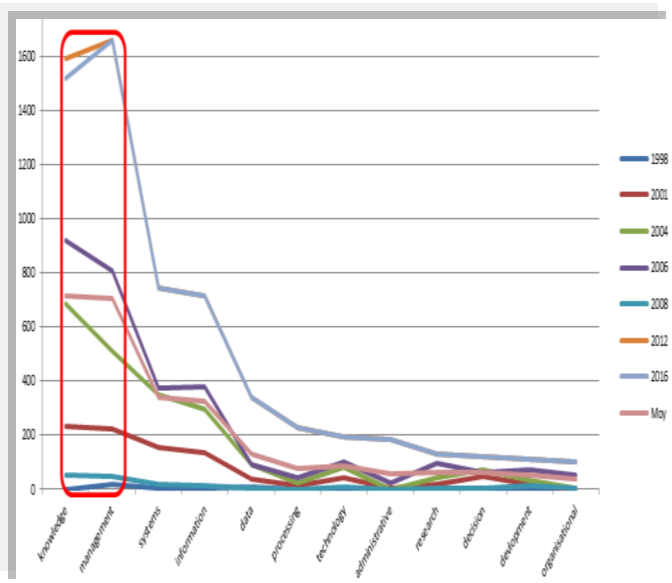

Figure 5. Densities of authors' keywords of KM, per year

In all years, the word **management** is less dense than the word **knowledge**, because there are other stages in KM process (e.g. knowledge transfer, knowledge capitalization…etc.).

TABLE XI DENSEST WORDS IN AUTHORS' KEYWORDS OF KM

Dense word	Avg	Dense word	Avg	Dense word	Avg
Knowledge	714	Data	131	Research	60
management	703	Technology	88	administrative	57
Systems	341	Processing	76	development	52
information	**323**	Decision	62	organisational	38

The results of the density in the authors' keywords are similar to the density in titles (TABLE XI), because in general the authors get their keywords from the title. Unlike IC, KM has no common dense words in the titles and authors' keywords at least in these first rows of TABLE XI except **information**, **knowledge** and **management**.

D. *Analysis of Zipf's law*

The analysis of Zipf's law is different from the densest words analysis seen in Section C. To apply Zipf's law, all words are taken into account, including the stopwords.

1) Analysis of Zipf's law in IC: A set of 655 among 1500 full articles in IC are used to create a corpus on which Zipf's law was applied. A set of 21365 different words appears in the corpus with different frequencies. The existing Zipf's law represents the found frequencies of words and the theoretic Zipf's law represents the theoretic frequency of word=Size of the corpus of text with stopwords (6911590 words) divided by the rank of the word after ordering by descending order of existing frequencies.

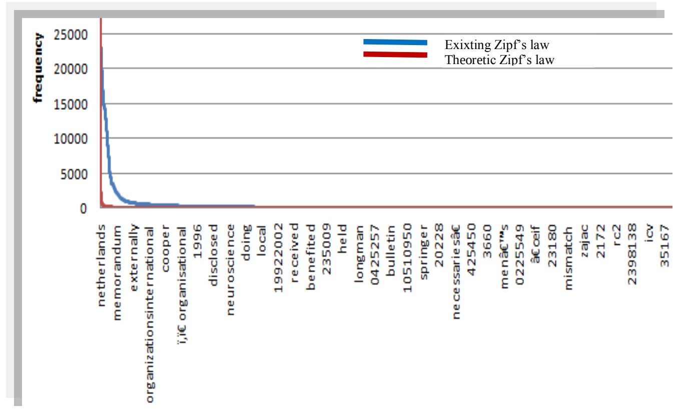

Figure 6. Zipf's law in IC corpus

Fig. 6 demonstrates that Zipf's law is globally verified and slightly different in the middle of curve.

2) Analysis of Zipf's law in KM: A set of 2420 among 10000 articles in KM are used to create a corpus on which we applied Zipf's law. A set of 18466 different words appears in the corpus with different frequencies. The theoretic Zipf's law represents the theoretic frequency of word=Size of the corpus of text with stopwords (6470344 words) divided by to the rank of the word after ordering by descending order of existing frequencies.

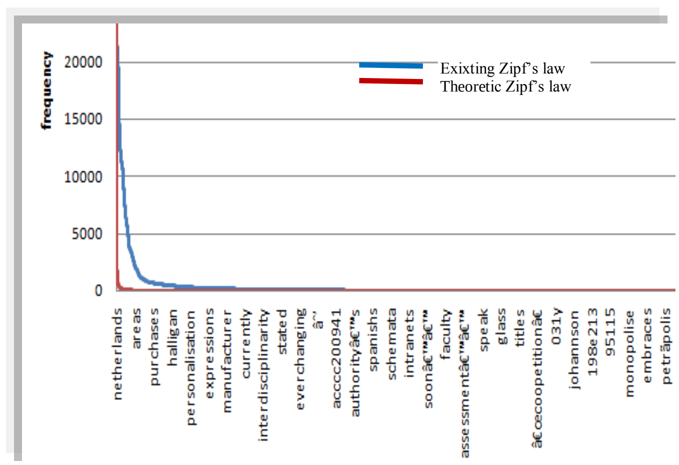

Figure 7. Zipf's law in KM corpus

Fig. 7 demonstrates that Zipf's law applied in KM is globally less verified than IC and slightly different in the middle of curve.

V. CONCLUSION

In this work, scientific articles (abstracts and full articles) were quantitatively analyzed to draw qualitative conclusions. These articles are related to the areas of intellectual capital and knowledge management. Each area was analyzed separately and it has been found different results especially for dense words. The densities are closely related to the scientific production of the area.

Keyword analysis allows the identification of existing levels of meaning in scientific articles as well as the unexpected trends in the area. Eight programs have been written in R language and applied to IC and KM. The goal is to delve into the exported data files from SCOPUS to collect the authors' keywords of scientific articles and calculate their frequencies. The full list of authors' keywords in both areas prepares the ground for the area glossary and ontology design. The goal was to find the frequencies of authors' keywords in the articles, keywords proposed by a user as needed and densest words in several items of articles (titles, abstracts, full files and authors' keywords). The latter is the most significant; it allows discovering new concepts introduced in the area for enrichment of glossaries and ontologies which could be proposed for a research area. KM and IC have many common concepts as management, performance, model, value.

Generally, Zipf's law is valid in IC and KM fields for the database SCOPUS. Despite the fact that the sample of full articles in KM is larger than IC ones, the corpus of IC is richer in keywords.

As perspectives, we could apply other sets of keywords indicators to draw other conclusions. This work prepares the ground for the creation of glossaries and ontologies for both IC and KM areas and it addresses the semantic relationships between keywords.

For the representation of frequency, one would think to develop a tool for representing words in clouds as the frequency and relevance criteria specified by the user.

For authors' keywords, an analysis by group of words is possible because there are synonyms which may offer a sum of frequencies and one can calculate the frequency of the group of words. This grouping step requires a thorough knowledge of words and their use in the search area, and if we would do it automatically, this would require an additional effort of semantic links expressed by other ontologies or semantic analysis tools. The extension can be the logical combinations between several keywords in the search query. At the end, the designed programs can be applied to any area of research exported from SCOPUS database. To use them for other databases, it will be necessary to slightly adapt them to the shape of input data.

REFERENCES

[1] M. Karl Wiig, "Integrating intellectual capital and knowledge management", Long Range Planning, vol. 30, No. 3, pp. 399-405, 1997.

[2] I. J. Lopez-Veyna, J. V. Sosa-Sosa and I. Lopez-Arevalo, "A low redundancy strategy for keyword search in structured and semi-structured data Information Sciences", Information Sciences, vol. 288, pp. 135-152, 2014.

[3] C. C. Bang, Information systems frontiers: Keyword analysis and classification, Information Systems Frontiers, vol. 17, No. 1, pp. 217-237, 2015.

[4] C. Jaewoo and K. Woonsun, "Themes and trends in Korean educational technology research: A social network analysis of keywords", Procedia - Social and Behavioral Sciences, vol. 131, pp. 171 – 176, 2014.

[5] W. Bihu, X. Honggen, D. Xiaoli, W. Mu and X. Lan, "Tourism knowledge domains: A Keyword Analysis", Asia Pacific Journal of Tourism Research, vol. 17, No. 4, pp. 355-380, 2012.

[6] G. F. Khan and J. Wood, "Information technology management domain: emerging themes and keyword analysis", Scientometrics, 105: 959. doi:10.1007/s11192-015-1712-5, 2015.

[7] A. Andrés, Measuring academic research: How to undertake a bibliometric study, Chandos Publishing, 2009.

[8] W. Li, "Random Texts Exhibit Zipfs-Law-Like Word Frequency Distribution", IEEE Transactions on information theory, vol. 38, No. 6 pp. 1842-1845, 1992.

[9] L. A. Adamic and B. A. Huberman, "Zipf's law and the Internet", Glottometrics, vol. 3, pp.143-150, 2002.

[14] M. Habibi and A. Popescu-Belis, "Diverse keyword Extraction from Conversations", The fifty first Annual Meeting of the Association for Computational Linguistics, pp. 651-657, 2013.

[15] D. Trieschnigg, D. Nguyen and M. Theune, "Learning to Extract Folktale Keywords", University of Twente, pp 1-9, 2013.

[16] L. Marujo et al, Automatic Keyword extraction on Twitter, University of Carnegie Mellon, 2014.

[17] S. Klapdor, M. Anderl, F. Wangenheim and J. H. Schuman, "Finding the right words: the influence of keyword characteristics on performance of paid search campaigns", Journal of Interactive Marketing, vol. 28, pp. 285–301, 2014.

[18] R. Rizzo, M. José and M. Pérez, "A key perspective on specialized lexis: keywords in telecommunication engineering for CLIL", Procedia - Social and Behavioral Sciences, vol. 198, pp. 386 – 396, 2015.

[19] A. Onan, S. Korukoglu and H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification", Expert systems with applications, vol. 57, pp. 232-247, 2016.

[20] I. Tuomi, "Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory", Journal of Management Information Systems, vol. 16, No. 3, pp 103-117, 1999.

[21] S. Aitouche, M. D. Mouss, A. Kaanit, K. Bouhafna, W. Mammeri and A. Chabbi, "SKACICM a method for development of knowledge management and innovation system e-KnowSphere", International journal of Knowledge and Web Intelligence (IJKWI), vol. 5, No. 2, pp. 105-126, 2015.

[22] S. Aitouche et al, "A hybrid method to develop a knowledge management system", The Sixth International Conference on Information, Process, and Knowledge Management (EKNOW 2014), IARIA, pp. 77-83, Mar 2014, ISSN: 2308-4375 ISBN: 978-1-61208-329-2,

[23] M. Titah, M.D. Mouss, S. Aitouche, "An implementation tool for the expertise model using CommonKADS methodology", The Sixth International Conference on Information, Process, and Knowledge Management (EKNOW 2014), IARIA, pp. 70-76, Mar 2014,ISSN:2308-4375, ISBN: 978-1-61208-329-2.

[24] H. Barki, S. Rivard and J. Talbot, "An information systems keyword classification scheme", MIS Quarterly, vol. 12, No. 2, pp. 299-322, 1988.

 ISBN: 978-1-61208-542-5