Webology (ISSN: 1735-188X)
Volume 19, Number 6, 2022

Keyword Analysis In Articles By Frequency Of Occurrence In The Text

Albert Miyer Suarez Castrillón[1], Sir-Alexci Suarez Castrillon[2], Marta Milena Peñaranda Peñaranda[3]

[1] Faculty of Engineering and Architecture, GIMUP Research Group. Universidad de Pamplona, Colombia.

[2] Faculty of Engineering, GRUCITE Research Group, University Francisco de Paula Santander Ocaña, Colombia

[3] Master in Organizational Management. Researcher of the GIDSE group. University Francisco de Paula Santander Ocaña, Colombia.

ABSTRACT

This research analyzes the relationship between the frequencies that a keyword appears in the publication, as a reference for the creation of the appropriate metadata of the article. The frequency is taken into account and the weight is analyzed according to the order in which the frequency is higher, as well as in the areas of Mechanical and Electrical Engineering, Business Administration and Accounting and Civil Engineering. The results show that according to the publication, the use of keywords may change, and may affect the search of the area according to its metadata and that the area may not relate the keywords with their frequency of repetition.

Keywords: word frequency, scientific articles, metadata, Norte de Santander.

1. INTRODUCTION

Metadata helps to retrieve information in a search and can be an important factor in getting news, publications or product sales to the end consumer. Articles that are published in different ways in internal academic publications can make use of keywords so that metadata becomes a tool to help in the distribution of content.

For this reason, an analysis of keywords and frequency for publications in the Norte de Santander Region is carried out. For the selection of the publications, the area of each one is taken into account, three articles per publication are randomly analyzed, since the quality of each one is already verified, having been reviewed by reviewers, and the volume, year and number are also randomly selected. Each one is assigned a code for subsequent data analysis, the publications are from the University of Pamplona with the journal of Advanced Technologies (RCTA) for the area of mechanical and electrical engineering, Universidad Francisco de Paula Santander Ocaña (UFPSO), with the journal Ingenio for the area of civil engineering and the journal Profundidad for the area of Administration and Accounting.

For RCTA the first article focuses on the experimental farm of the UFPSO where lighting control systems are needed for the poultry area (Villegas et al., 2020)-(A1-1), which can determine the appropriate parameters to configure natural light and artificial light. Simulation is a tool that allows to anticipate the realization to know how the final result of a part or process would be obtained (Ortiz et al., 2020)-(A1-2), and saving time and money, in this case a multidimensional simulation is performed for a combustion compressor by heat transfer. Finally, the open vehicle routing problem is analyzed using the "Wait and See" solution (Castaño et al., 2020)-(A1-3). A methodology for the prediction from ground vibrations due to blasting is presented and recommendations are given with criteria to evaluate them (Oliva-González & Fort-Villavicencio, 2019)-(A2-1). The paving of regional areas is of vital importance for the economy in Colombia, and the materials must be appropriate, a study of the asphalt mixture "MAPIA", characterizes this material, its installation process and properties for the improvement of a section in the municipality of San Martin in the Department of Cesar (Gómez-Galván et al., 2019)-(A2-2). Continuing with the previous context in the city of Barrancabermeja there is an abandoned railway corridor and it crosses the entire urban case of the city (Estupiñán et al., 2019)-(A2-3), while the automobile hull increased by 50% due to the number of existing motorcycles, that is why a simulation is performed to demonstrate the benefit that would be achieved by enabling this passage as a vehicular road. The pandemic brought changes in the way of living of the world population, within it the change in consumption was one of the most notable, and each region has its particularities, as it is in the metropolitan area of Bucaramanga, where it is demonstrated that the increases are given in products of the family basket, cleaning and cleaning, while decreases in the consumption of alcohol, pets, footwear or textiles (Cuentas & Barajas, 2021)-(A3-1). The strengthening of the enterprise and its administrative management can help to improve the production and commercialization of Wayuu handicraft products (Orozco et al., 2021)-(A3-2). The in Profundidad publication in the third article investigates the possibility of generating a patent through technical capabilities that help to improve the production process of panela, which is being carried out in an artisanal manner (Cely & Zamudio, 2021)-(A3-3).

2. METHODOLOGY

Relevant scientific publications from the department of Norte de Santander in different areas such as: Business Administration, Mechanical Engineering and Civil Engineering are analyzed in order to determine if the key words are extracted from the frequency with which they appear in the article. Table 1 describes the parameters of year of publication, volume, number, journal and area where each article is selected. The articles are randomly selected in each area and journal, since they have all been reviewed by peer reviewers and have a recognized scientific quality.

Table 1. Publication selection, volume and journal area.

Year	Volume	Number	Publicatión	Área
2020	3	Special edition	RCTA	Mechanical and Electrical Engineering
2019	16	1	Ingenio	Civil Engineering
2021	15	15	Profundidad	Business administration and accounting

The frequency of each word in the publication is calculated by extracting all the words and grouping them in 3 categories, if they are within the first 25 words a weight of 10 is given according to the number of words, if they are within 26 and 100 words a weight of 6 is given, and if they are above the first 100 words a weight of 1 is given. The codes are assigned with the letter A followed by a number that indicates the order of the publication as previously indicated, and after the hyphen the order of the article with a number as it appears in the introduction.

Calculations are made using the following equations. Where V=score frequency per item

$$\mathrm{V} = \frac{100 * \text{keyword}}{\text{total keywords}} \qquad (1)$$

And the article value and frequency of publication is given by equations 2 to 5. Where P is equal to the value of the score per article, according to the weight in the publication. And where FP is the final frequency per publication of the 3 articles, with their respective codes assigned for each publication.

$$\mathrm{P} = \mathrm{V} * 10 \qquad (2)$$

$$\mathrm{P} = \mathrm{V} * 6 \qquad (3)$$

$$\mathrm{P} = \mathrm{V} * 1 \qquad (4)$$

$$\mathrm{FP} = \frac{\mathrm{P(Code\ 1)} + \mathrm{P(Code\ 2)} + \mathrm{P(Code\ 3)}}{3} \qquad (5)$$

3. RESULTS

Table 2 shows that for the article with code A1-1, all the reserved words are used more frequently in the text of the article, highlighting words such as PID and PLC, as well as the Matlab programming language. In code A1-2 the trend is maintained and all the keywords are used more frequently in the text of the research, in total 771 unique words are analyzed, with 4 keywords. Code A1-3, has 1163 words in the article and 5 keywords are among the most cited, and only one presents a medium value (Table 2). The FP value in the 3 articles is quite high with 9.77%.

Table 2. Frequency of code A1 items.

Code	keyword	Among the first 25	Between 26 and 100	Over 100
A1-1	Closed Loop Control	X		
	MatLab	X		
	Transfer function,	X		
	PID	X		
	Programmable Logic Control (PLC)	X		
A1-2	Ignition-compression engine	X		

	Conjugate heat transfer	X		
	Combustion	X		
	Simulation	X		
A1-3	Stochastic demand	X		
	Stability		X	
	2-point estimation method	X		
	Open routing	X		
	Wait and see solution	X		
	Vehicle routing	X		

Article A2-1 complies in citations, since all words are pronounced in the algorithm. A2-2 have a total of 1535 words analyzed, and with 5 keywords, where none is in the highest frequency range, likewise for article A2-3, with 5 reserved words and with low frequency in the text (Table 3).

If the results of the 3 articles are analyzed, the keywords represent a FP of 4, which is quite low within the text of the articles.

Table 3. Frequency of items code A2.

Code	**keyword**	**Among the first 25**	**Between 26 and 100**	**Over 100**
A2-1	Structural damage	X		
	prediction methodology	X		
	ground vibrations	X		
	blasting	X		
A2-2	Natural asphalt			X
	Pavement structure			X
	MAPIA			X
	Asphalt mix			X
	Cold mix			X
A2-3	Transport models			X
	Mobility			X
	Railway track			X
	Simulation			X
	Streets			X

Article A3-1, presents a title and keywords related to each other, however in the body of the article does not include as such these words and focuses more on the business part as such. Article A3-2 has 1457 unique words, and the conclusions is that the reserved words are within the highest levels, which complies for possible citation of the topic in search engines. The article A3-3 has an extension of 1919 unique words, and 2 of the keywords are within the highest frequency, but a word as important as patent, which may be the final result of the research is not within the first 100 words, which denotes one more word in the future, because the process for obtaining the patent is not shown. While panela appears 85 times and technological watch 24 times (Table 4).

The frequency of the keywords in the 3 articles represents a FP value of 5.47 as mean, and it is a value where it does not make much reference to the keywords within the articles, this is also due to the random way of the selection of each article, because article A3-2 and A3-3 present high results, and A3-1 does not use much repetition or the keywords are named frequently in the research.

Table 4. Frequency of code A3 items.

Code	**keyword**	**Among the first 25**	**Between 26 and 100**	**Over 100**
A3-1	Consumption behavior			X
	Covid-19			X
	Trends			X
A3-2	Administrative	X		
	Craftswomen		X	
	Entrepreneurship	X		
	Strengthening		X	
	Organizational	X		
	Wayuu	X		
A3-3	Technological Development		X	
	Patent			X
	Production of panela	X		
	Technology watch.	X		

In Figure 1, it can be seen that articles with code A1 represent the highest frequency of words in the article, while code A2 does not make frequent reference and those with code A3 have an average frequency per article.

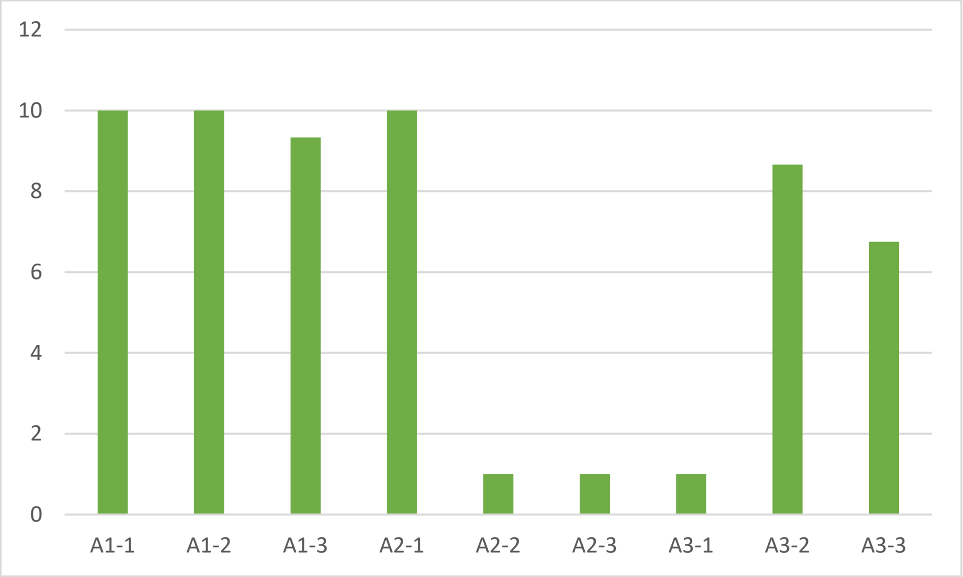

Figure 1. Frequency value per item by weight.

4. CONCLUSIONS

The relationship between the selection of key words in an article and its content can be denoted by the frequency in which they appear in the text. The analysis of publications from the Norte de Santander region was carried out following a random selection of articles, which may be an error because the sample is small, but if they have a similarity in the results, it may serve as a sample of their characteristics. The publications have shown that keywords are the most frequently used in the article as in RCTA and Profundidad, but in Ingenio they are not used as frequently.

REFERENCES

Castaño, A. O., Ocampo, E. M. T., & Rendón, R. A. G. (2020). Solución "wait and see" para el problema de ruteo abierto de vehículos con demandas estocásticas bajo un esquema de estimación por puntos. Revista colombiana de tecnologias de avanzada (RCTA), 3(Especial), Art. Especial. https://doi.org/10.24054/16927257.vEspecial.nEspecial.2020.851

Cely, S. R. H.-, & Zamudio, M. T.-. (2021). Capacidades y tendencias tecnológicas en el proceso de producción de panela artesanal. Un estudio de vigilancia tecnológica. Revista Científica Profundidad Construyendo Futuro, 15(15), Art. 15. https://doi.org/10.22463/24221783.3310

Cuentas, N. C. P.-, & Barajas, C. E. T.-. (2021). Comportamiento de consumo a raíz de la pandemia en Bucaramanga y AMB. Revista Científica Profundidad Construyendo Futuro, 15(15), Art. 15. https://doi.org/10.22463/24221783.3246

Estupiñán, Y. F. M.-, Martínez-Guerra, C., & Carrero-Monroy, O. (2019). De vías férreas a carreteras urbanas. Análisis para la ciudad de Barrancabermeja. Revista Ingenio, 16(1), Art. 1. https://doi.org/10.22463/2011642X.2345

Gómez-Galván, M., Gallardo-Amaya, R., & Macgregor-Torrado, A. A. (2019). Pavimentación con asfalto natural MAPIA. Estudio de caso: Proyecto mejoramiento de la vía El Diviso – Torcoroma del municipio de San Martin, Cesar. Revista Ingenio, 16(1), Art. 1. https://doi.org/10.22463/2011642X.2334

Oliva-González, A. O., & Fort-Villavicencio, R. (2019). Metodología para la predicción de las vibraciones del terreno inducidas por voladuras y sus efectos en las estructuras. Aplicación en un caso real. Revista Ingenio, 16(1), Art. 1. https://doi.org/10.22463/2011642X.2381

Orozco, J. J. L.-, Camargo, J. C. R.-, & González, R. A.-. (2021). Implicaciones del Emprendimiento en el Fortalecimiento de la Gestión Administrativa y Organizacional de las Artesanas Wayuu. Revista Científica Profundidad Construyendo Futuro, 15(15), Art. 15. https://doi.org/10.22463/24221783.3249

Ortiz, Y., Flórez, E. G., & Laguado, R. I. (2020). Simulacion multidimencional de transferencia de calor en un compresor de combustion interna en el encendido. Revista colombiana de tecnologias de avanzada (RCTA), 3(Especial), Art. Especial.

Villegas, J. F. P., Murcia, F. M., & Camperos, J. A. G. (2020). Sistema de control de iluminación para los galpones avícolas en la granja experimental de la universidad francisco de paula santander – ocaña (UFPSO). Revista colombiana de tecnologias de avanzada (RCTA), 3(Especial), Art. Especial. https://doi.org/10.24054/16927257.vEspecial.nEspecial.2020.848