Determination Of The Measurement Factor In Citations Of Educational Publications Using Natural Language Processing

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

[1] *Faculty of Engineering, GRUCITE Research Group, Universidad Francisco de Paula Santander, Ocaña, Colombia*
[2] *Administrative and Economic Sciences Faculty, Universidad Francisco de Paula Santander, Ocaña, Colombia*
[3] *Faculty of Engineering and Architecture, University of Pamplona, Pamplona, Colombia*

Abstract

An investigation is presented through the processing of natural language and text analysis found in the abstracts of the educational publication Profundidad of the Faculty of Administrative and Economic Sciences at the Universidad Francisco de Paula Santander Ocaña. The indicators analysed are magnitude, year of publication, importance weighting and sentiment index, so that it can be determined if the sentiment index determines the percentage of citations of the articles in importance with reference to the other indicators. The number of articles in the publication is 107 articles and a sample of 13 abstracts selected among those with the highest index and similar in score. The analysis is performed using the words in Spanish, the official language of the publication, since the translation may vary the score due to its meaning.

Keywords: summary journal; natural language; sentiment index, publication depth.

1. Introduction

The importance for a researcher to read a complete article can derive from the abstract and its composition, following guidelines such as the precision and conciseness of the writing (Tullu, 2019), and the length usually comprises about 250 words (Sirisilla, 2022). In itself the abstract helps the reader to understand the research, to know the limitations and findings and to classify the topics of each article for search engines and academic databases, even when the articles are paid the abstracts are available to give the reader an idea if it is good to pay for the full article, so it simply becomes the cover letter of the researcher (Camps, 2010). The H-index of an article measures the number of citations and is an indicator that can determine the quality of the article (Scholar academic, 2022), although there is no lack of bad practices that can distort the value of the index. Based on the above, the article abstract can determine the number of citations it obtains by increasing the index, however to measure the quality of the abstract new tools have been created based on natural language processing and analysis of the text that comprises it to determine whether an abstract maintains a positive or negative index and to know if it can influence

the reader (Umer et al., 2021). Text analysis to measure sentiment is largely applied in social networks such as twitter (Al Baghal et al., 2021; Kharde & Sonawane, 2016) or facebook (Albesta et al., 2021), and can reflect the opinions of customers or users. A positive index can indicate that things or products are done well and the negative index will help to improve the services provided by the companies.

To carry out this research, 13 abstracts of the most referenced articles are analyzed according to the H index of google academic, which have more tendency to be cited in the faculty of Administrative and Economic Sciences at the Universidad Francisco de Paula Santander Ocaña (UFPSO), starting with sustainability in agriculture (Fonseca-Carreño, 2019), where the productive and marketing chain of the Rubus Glaucus in the Sumapaz is discussed; subsequently the importance of financial management (Zambrano-Mercado et al., 2020) in territorial development plans to execute efficient projects and programs; then a bibliometric study (Manzano-Durán et al., 2021) on sustainable projects is carried out after analyzing more than 550 documents in the Scopus database, where it is found that more research should be published in book format.

Another research analyzes the culture of entrepreneurship (Pineda-López & Manzano-Sanguino, 2018) with the aim of creating an entrepreneurial spirit in business administration students and thus channeling entrepreneurial skills and abilities; also at the level of Ocaña, a research is conducted to know the level of service provided by clinical laboratories (Coronel-Nuñez et al., 2019), some of which have more than 46 years of creation, finding that many do not have adequate capacity to compete in the market; a study on migration of Venezuelan migrants in Barranquilla (Salas-Hernandez et al., 2019), where the contribution and strengthening of their economic activities help substantial growth of the city and the country.

A study on B companies (Rodríguez-Téllez & Pacheco-Sánchez, 2019) to know the socially responsible companies, which determines that the company should not only serve the customer, but also society, finding that there was no certification on the model in the city of Ocaña; a research on augmented reality and the importance in teaching in financial mathematics is also presented, allowing to be applied to learning objects through a mobile application, which is intended to be a technological aid tool that can be used by means of financial mathematics students (Muñoz-Hernandez et al. , 2020).

Research on organizational strategies that help collective welfare in companies is also analyzed (Ortiz-Gutiérrez et al., 2021); a reading culture is also presented as an essential activity in people (Moreno-Mendoza & Coronel-Arenas, 2018), looking for habits that prevent the adoption of reading, which contributes to the acquisition of new knowledge and formation of professional skills; socio-formative research is also analyzed in such a way that the real experience with the theoretical one helps to form better graduates (Cruz-Vega, 2019); the last two researches talk about triadic thinking and sociodemographic variables of business administration students at UFPSO (Lemus-Quintero, 2021) and the last one focuses on the psychology of color which can influence advertising services, which through different strategies can influence the final service (Barbosa & Sanchez, 2017).

The objective of the research is to analyze the sentiment index, to determine if the higher index of a positive summary can determine that an article is more cited than another, also analyzing the variables of dimension, year of publication and weighting of importance on the publication depth of the Faculty of Administrative and Economic Sciences in the UFPSO.

2. Method

The academic profile of the journal Profundidad and its H index obtained from Google Scholar are used to extract the number of articles published in the magazine Profundidad, so there are a total of 107 articles and 13 articles with an H index between 11 and 4 are taken as a sample, the field to be analyzed

is the Abstract of each article and the number of years it has been published, The field to be analyzed is the Abstract of each article and the number of years it has been published, where the first articles with the highest H index are not taken into account, because the sample of these 13 articles has a similar index but varies in the years, which will allow knowing the impact that the Abstract can generate when cited and read by other researchers, in Figure 1 the selection of the articles is observed.

1) *Sustainability In Agroecological Family Agriculture: Castilla Blackberry In Sumapaz* (Fonseca-Carreño, 2019)	2) *Financial management in land management plans as a tool for urban development* (Zambrano-Mercado et al., 2020)	3) *Sustainability and sustainable projects: Bibliometric study* (Manzano-Durán et al., 2021)
4) *The Entrepreneurial Culture In Business Administration Students* (Pineda-López & Manzano-Sanguino, 2018)	5) Characterization of the clinical laboratories of the municipality of Ocaña (Coronel-Nuñez et al., 2019)	6) *Entrepreneurship And Informal Work Of Venezuelan Immigrants, Barranquilla Case 2015-2018* (Salas-Hernandez et al., 2019)
7) *Companies B: A Diagnosis On Its Applicability* (Rodríguez–Téllez & Pacheco-Sánchez, 2019)	8) *Augmented reality for education of financial mathematics. An application to improve university academic performance* (Muñoz-Hernandez et al., 2020)	9) *Organizational development theory: Organizational and labor inefficiency* (Ortiz-Gutiérrez et al., 2021)
10) *Adverse habits and barriers for the adaptation and adoption of reading* (Moreno-Mendoza & Coronel-Arenas, 2018)	11) *Socio-educational evaluation in the research processes of the children's pedagogy programme* (Cruz-Vega, 2019)	12) *Triadic thinking and its relationship with sociodemographic variables of Business Administration students* (Lemus-Quintero, 2021)
	13Color Psychology: Strategies immersed in the service of advertising agencies (Barbosa & Sánchez, 2017)	

Figure 1. Sample obtained from the profile of the journal Depth in Google Scholar.

A total of 1958 words are analyzed, which are extracted to form vectors composed of one token per word. The total number of tokens per article can be seen in Figure 2. However, since tokens can distort the information, the complete sentences in each Abstract are analyzed to determine whether the message is positive.
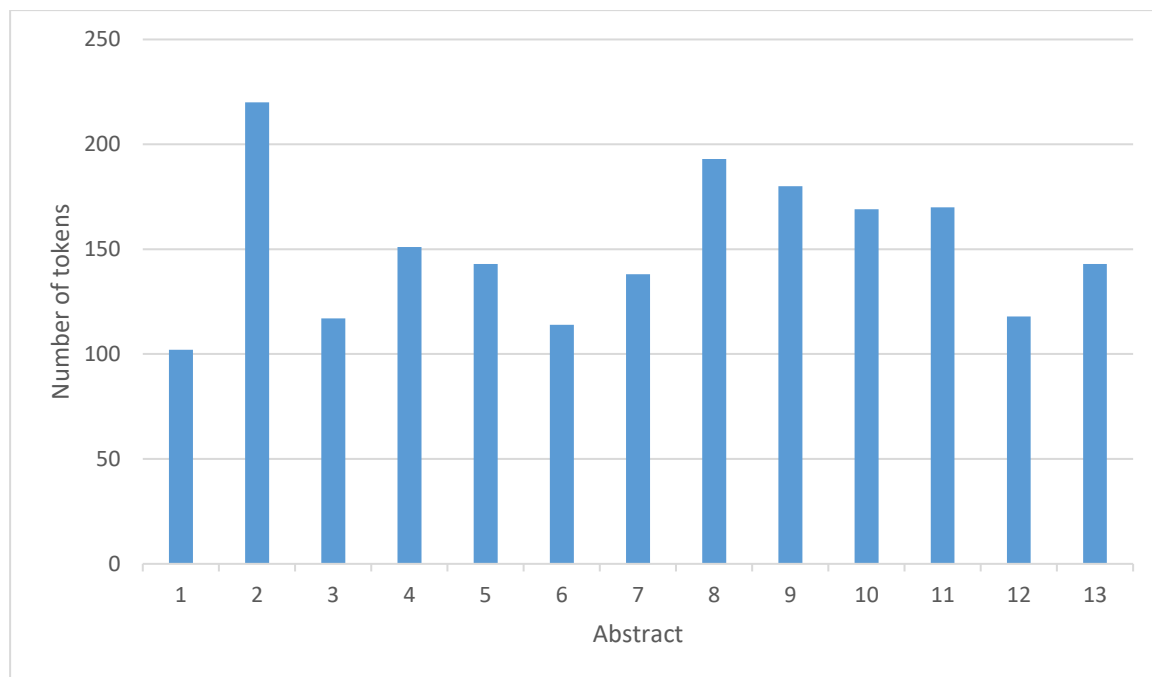
Figure 2. Number of tokens per abstract.

To determine the positive, neutral or negative index the google Natural Language API is used, where the values have a range from 1 very positive to -1 very negative, the full range of index scores is determined as follows:

0.25 — 1.0	-0.25 — 0.25	-1.0 — -0.25

Where:

Positive: 0.25 to 1.0.

Neutral: -0.25 to 0.25

Negative: -1.0 to -0.25

The publications are in Spanish, the official language of the publication Depth, which ensures that translation into another language does not change the weighting of the sentiment index. To determine the importance of the score and its index in each abstract, a weighting will be taken with the elapsed time of publication, which does not exceed 5 years, since this is the range where most citations are obtained, allowing to reduce the importance because the more years an article has been published, the more likely it is to have been cited, so that the importance score is measured by the weighting equations in Table 1.

Table 1. Weighting equations

Years elapsed	Index discount	Equation	Weighting
1	0%	(1)	$P = \sum_{i=1}^{13} indice_i$
2	20%	(2)	$P = \sum_{i=1}^{13} indice_i * 0.8$

| 3	30%	(3)	$P = \sum_{i=1}^{13} indice_i * 0.6$
4	40%	(4)	$P = \sum_{i=1}^{13} indice_i * 0.4$
5	50%	(5)	$P = \sum_{i=1}^{13} indice_i * 0.2$

3. Results

From the 13 abstracts of each article, the number of sentences present in each one is extracted in such a way that the positive index of the message is not distorted, since a single token sometimes does not contain all the relevant information. Based on this, 46 sentences are obtained divided into the different abstracts as shown in Figure 3, and the number of tokens per sentence is also measured in order to know whether the higher or lower number affects the index of the message. Table 2 shows in green color the quantity with smaller tokens of each sentence and in purple color the larger quantity of the sentence in each abstract. It can be deduced that the higher or lower number of tokens does not affect the number of citations because abstract 6 has only one sentence and is well placed within the citations, and it is not relevant whether the first sentence is higher or lower in number of tokens, nor is the number of tokens significant in any location. It is also observed that the referential mean of the sentences per abstract has a value of 4, and they are among the most cited.

Table 2. Number of tokens per sentence in each item.

Abstract	Vector of sentences					
	1	2	3	4	5	6
1	27	29	12	34		
2	50	31	29	17	31	62
3	36	29	31	21		
4	68	83				
5	28	41	74			
6	114					
7	39	45	15	39		
8	33	83	24	53		
9	58	28	94			
10	63	51	28	27		
11	75	51	44			
12	32	23	38	25		
13	25	35	38	45		

Menor ▮   Mayor ▮
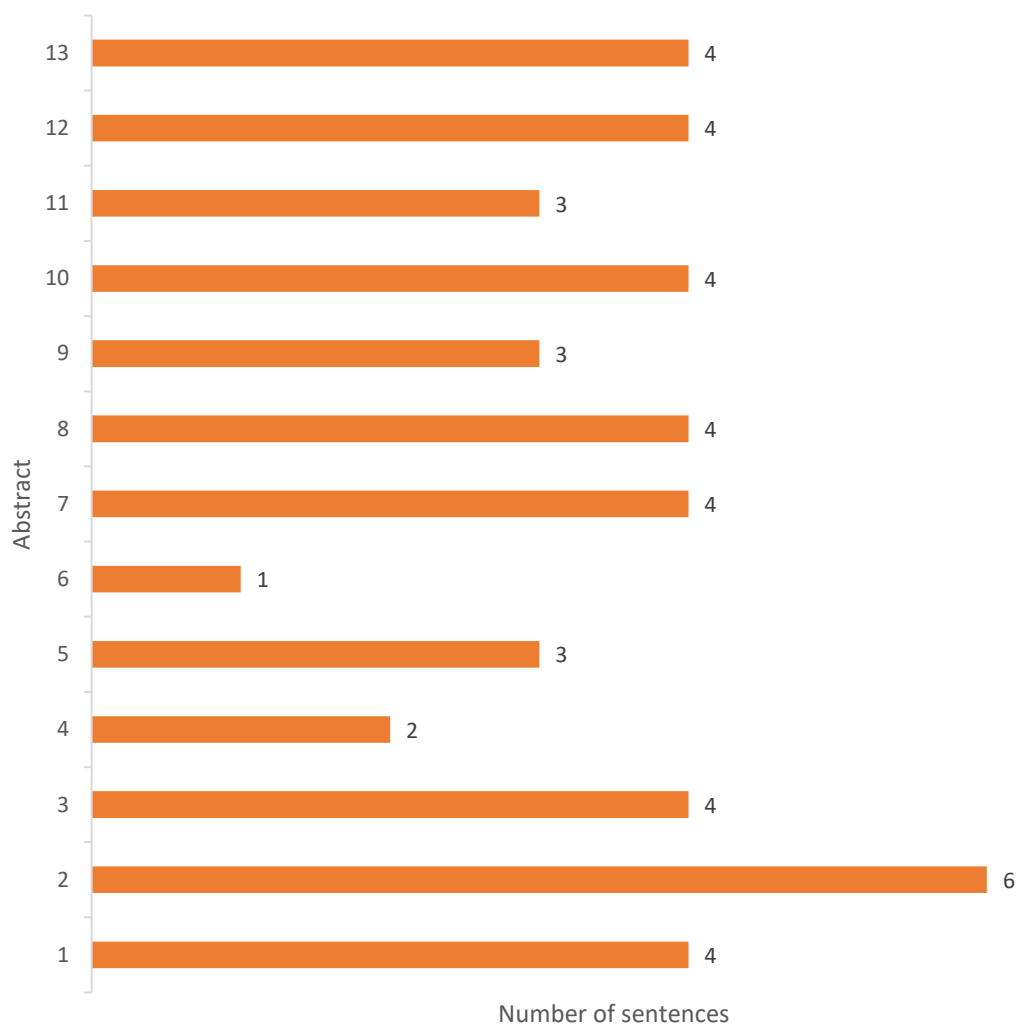
Figure 3. Number of sentences per Abstract number.

Table 3 shows the classification according to the score obtained by each sentence that composes an abstract, where the number of positive or neutral sentences does not condition the final result of the abstract, and in general all the sentences that compose each abstract tend to be positive or at least neutral, which reflects that the research is composed of research with relevant results, which can help the contribution of science, although it is true that the positive score tends to be more neutral, which could denote that the research has fields to grow in the future or at least continue to make contributions. It is also reflected that the first sentence of each abstract begins positively with 76.92%.

Table 3. Sentiment score by phrases.

	Vector of phrases					
Abstracts	1	2	3	4	5	6
1	0,4	0,5	0,4	-0,1		
2	0,4	0,4	0,1	0,1	0,5	0,2
3	0,3	0,2	0	0,4		
4	0,6	0,2				
5	0,1	0,1	0			
6	0,3					

|					
7	0,4	0	0,2	0,2	
8	0,7	0,1	0,3	0,1	
9	0,5	0,3	0,5		
10	0,9	0	0,2	0,3	
11	0,1	0,4	0,4		
12	0,2	0,2	0,6	0,1	
13	0,3	0,8	0,3	0	

Positive ▮   Neutral ▮

Table 4 shows the total results of each abstract, where the score is positive in 76.92%, with discrete values of 0.3 in the majority, and neutral in 23.08%. While the magnitude does vary in its intensity, even in the abstrats with neutral index, for example a neutral index of asbtract 3 has a magnitude of 1.1 higher than a positive one such as asbtract 6 with 0.3. There is also no relationship between the positive and neutral index of the abstracts with the number of citations, not even with the size in the number of sentences, therefore for the region and academic analysis of the university they do not take it into account.

But if we analyze the years that have elapsed since they were published, we can see that the period is a maximum of 5 years, which means that for the journal depth is not so relevant the area of knowledge of the article or the positive index of the abstract, and its focus is on how long it was published.

Table 4. Sentiment score per abstract.

	Number Abstract												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Index Sentiment	0.3	0.3	0.2	0.4	0	0.3	0.2	0.3	0.4	0.3	0.3	0.3	0.3
Magnitude	1.5	2	1.1	0.9	0.3	0.3	1	1.4	1.4	1.5	1	1.2	1.5
Index h	11	10	9	8	8	7	6	6	6	5	4	4	4
Years	3	2	1	4	3	3	3	2	1	4	3	1	5

However, it is noted that articles published 1 year after publication have a higher H-index than articles published 5 years after publication, which is why the weight formula from equation 1 to 5 is applied (Figure 4). The results show that the abstracts with neutral index tend to decline in placement, and that in recent years the authors who publish in the journal do so in a more positive way with their findings in each research and it is observed that abstracts 3, 9 and 12 are placed with a higher weighting of importance only for being published in the last year (Table 5).
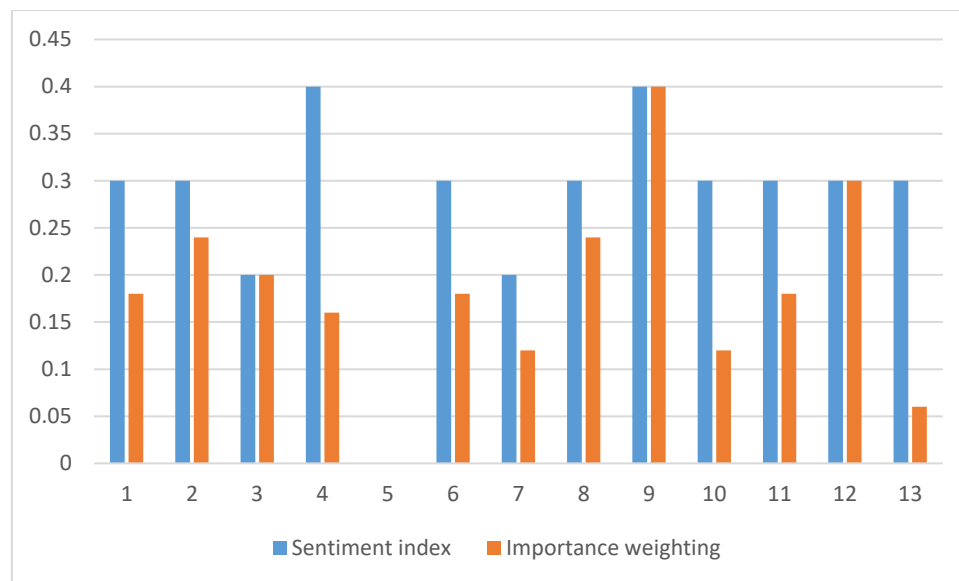
Figure 4. Importance weighting.

Table 5. Location according to importance weighting.

Abstract Number	1	2	3	4	5	6	7	8	9	10	11	12	13
Sentiment Index	0.3	0.3	0.2	0.4	0	0.3	0.2	0.3	0.4	0.3	0.3	0.3	0.3
Abstract Number	9	12	3	2	8	1	6	11	4	7	10	5	13
Weighting	0.4	0.3	0.2	0.24	0.24	0.18	0.18	0.18	0.16	0.12	0.12	0	0.06

4. Conclusions

The importance of abstracts in scientific publications can result in a higher citation of the publication, if its content is clear, concise and describes the findings and limitations, however, through text analysis it can be determined whether the positive, neutral or negative sentiment index influences citations. The results show that everything depends on the publication and the existing legislation based on the databases that governs the university, for this research it was found that the sentiment index and magnitude had no influence on citations, and the year was the factor that determined the highest number of citations, where the publication period is between 1 and 5 years old, and practically decreases considerably as the years go by.

References

Al Baghal, T., Wenz, A., Sloan, L., & Jessop, C. (2021). Linking Twitter and survey data: Asymmetry in quantity and its impact. *EPJ Data Science*, *10*(1), Art. 1. https://doi.org/10.1140/epjds/s13688-021-00286-7

Albesta, D. D., Jonathan, M. L., Jawad, M., Hardiawan, O., & Suhartono, D. (2021). The impact of sentiment analysis from user on Facebook to enhanced the service quality. *International Journal of*

Electrical and Computer Engineering (IJECE), *11*(4), Art. 4. https://doi.org/10.11591/ijece.v11i4.pp3424-3433

Barbosa, M. F. O., & Sánchez, C. A. P. (2017). Psicología del color: Estrategias inmersas en el servicio de las agencias de publicidad. *Revista Científica Profundidad Construyendo Futuro*, *7*(7), 39-45.

Camps, D. (2010). The Abstract: The letter of presentation for a scientific paper. *Colombia Médica*, *41*(1), 82-84.

Coronel-Nuñez, Y., Velasco-Burgos, B. M., & Cárdenas-García, M. (2019). Characterization of the clinical laboratories of the municipality of Ocaña. *Revista Científica Profundidad Construyendo Futuro*, *11*(11), 23-30.

Cruz-Vega, F. A. (2019). Evaluación socio formativa en los procesos investigativos del programa de pedagogía infantil. *Revista Científica Profundidad Construyendo Futuro*, *10*(10), 10-23.

Fonseca-Carreño, N. E. (2019). Sustentabilidad en la agricultura familiar agroecológica: Mora de Castilla en Sumapaz. *Revista Científica Profundidad Construyendo Futuro*, *11*(11), 12-22.

Kharde, V., & Sonawane, S. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. *International Journal of Computer Applications*, *139*, 5-15. https://doi.org/10.5120/ijca2016908625

Lemus-Quintero, J. A. (2021). Pensamiento tríadico y su relación con variables sociodemográficas de los estudiantes de Administración de Empresas. *Revista Científica Profundidad Construyendo Futuro*, *14*(14), 54-63.

Manzano-Durán, O., Peñaranda-Peñaranda, M. M., & Luna-Quintero, J. C. (2021). Sostenibilidad y proyectos sostenibles: Estudio bibliométrico. *Revista Científica Profundidad Construyendo Futuro*, *14*(14), 15-24.

Moreno-Mendoza, E. F., & Coronel-Arenas, M. F. (2018). Hábitos que impiden la adaptación y adopción de la lectura en estudiantes de Administración de Empresas. *Revista Científica Profundidad Construyendo Futuro*, *9*(9), 2-9.

Muñoz-Hernandez, H., Canabal-Guzmán, J. D., & Galarcio-Guevara, D. E. (2020). Realidad aumentada para la educación de matemática financiera. Una app para el mejoramiento del rendimiento académico universitario. *Revista Científica Profundidad Construyendo Futuro*, *12*(12), 37-44.

Ortiz-Gutiérrez, M. F., Galindo-Henao, A. M., Valbuena-Rodríguez, P. D., Clavijo-Peña, A. V., & Duarte-Leguizamo, J. D. (2021). Teoría del desarrollo organizacional: Ineficiencia laboral y organizativa. *Revista Científica Profundidad Construyendo Futuro*, *14*(14), 25-42.

Pineda-López, C. P., & Manzano-Sanguino, Y. D. (2018). La cultura emprendedora en los estudiantes de Administración de Empresas. *Revista Científica Profundidad Construyendo Futuro*, *8*(8), 2-9.

Rodríguez–Téllez, F., & Pacheco-Sánchez, C. A. (2019). Empresas B: Un diagnóstico sobre su aplicabilidad. *Revista Científica Profundidad Construyendo Futuro*, *10*(10), 2-9.

Salas-Hernandez, I. Z., Sagbini-Henriquez, H. S., & Salazar-Araujo, E. J. (2019). Emprendimiento y trabajo informal de migrantes venezonalas, caso Barranquilla 2015-2018. *Revista Científica Profundidad Construyendo Futuro*, *11*(11), 53-58.

Scholar academic. (2022). *español—Estadísticas de Google Académico*. https://scholar.google.com/citations?view_op=top_venues&vq=es

Sirisilla, S. (2022, septiembre 2). Abstract in Research Paper – Definition, Purpose, Types & Examples. *Enago Academy*. https://www.enago.com/academy/abstract-in-research-paper/

Tullu, M. S. (2019). Writing the title and abstract for a research paper: Being concise, precise, and meticulous is the key. *Saudi Journal of Anaesthesia*, *13*(Suppl 1), S12-S17. https://doi.org/10.4103/sja.SJA_685_18

Umer, M., Sadiq, S., Missen, M. M. S., Hameed, Z., Aslam, Z., Siddique, M. A., & Nappi, M. (2021). Scientific papers citation analysis using textual features and SMOTE resampling techniques. *Pattern Recognition Letters*, *150*, 250-257. https://doi.org/10.1016/j.patrec.2021.07.009

Zambrano-Mercado, R. A., Muñoz-Hernández, H., Brito-Aleman, C. C., & Caro-Flórez, A. P. (2020). Gestión financiera en planes de ordenamiento territorial como herramienta de desarrollo urbano. *Revista Científica Profundidad Construyendo Futuro*, *13*(13), 31-38.