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

Semantic Text Analysis To Determine The Positive Index Of The Ufpso Systems Engineering Program

Sir-Alexci Suárez Castrillón[1], July A. Gómez-Camperos[2], Haidee Y. Jaramillo[3]

[1]Department of Systems Engineering, Faculty of Engineering, GRUCITE Research Group, University Francisco de Paula Santander Ocaña, Colombia.

[2]Mechanical Engineering Department, Faculty of Engineering, GINSTI Research Group, University Francisco de Paula Santander Ocaña, Colombia .

[3]Department of Civil Engineering, Faculty of Engineering, GINSTI Research Group, University Francisco de Paula Santander Ocaña, Colombia.

ABSTRACT

This research presents a semantic analysis of text, based on the Systems Engineering program at the Universidad Francisco de Paula Santander Ocaña, with the purpose of knowing the index of positive mentions of the messages offered on the web page of the university and the respective program, and that due to online enrollment become the reference information for applicants to enter the program. Five factors were measured: Mission, Vision, Objectives, Professional Profile and Occupational Profile; obtaining a positive index between 55% and 65% in the previous factors. Wolfram software and the semantic text package were used to extract information, create tokens and classify the data.

Keywords: Semantic Analysis; UFPSO; professional profile; occupational profile; positive messages.

1. INTRODUCTION

Semantic text analysis allows extracting information from messages to determine the sentiment that may exist behind them (Alzboon, 2022; Setiawan, 2022). It is widely used through sentiment analysis where first the total number of words in a message is counted, then converted into n-grams or word frequency and by means of a classifier it is extracted whether the message is positive or negative, and the frequency with which some more relevant keywords may appear (Medagoda & Shanmuganathan, 2015).

At the scientific and publication level, work is being done on the relationship between the Abstract and the Keywords, because the words can be included by importance or by frequency in the text, with which the author's feelings can be analyzed (P. Zhang & Pan, 2020), it is clear that the Abstract can be the key with which a researcher can select whether or not to read the article. It also helps to know the opinions of researchers about the articles they read and if they

have similarity with scientific articles from different journals (Lin et al., 2022), where the mood or sentiment of the article can be known.

At the marketing level is where it is being used the most, because of the possibility that marketers have to know their customers, and based on this determine how they can correct their products or campaigns (Ng et al., 2022), within these aspects it is possible to monitor the changes of consumers about a product, which over time can change. One of the advantages is that it is possible to know the behavior of customers in times of decline in sales generated by crises such as COVID-19 (Cirqueira et al., 2020). If we talk about text messages, one of the social networks where sentiment processes are most performed is on Twitter, as for example in IKEA stores to know the emotion and opinion when opening a new store, where four cities were analyzed: Umea-Sweden, Halifax-Canada, Fishers-Ireland, and Sheffield-UK, being positive messages with very few negative messages, and that negative opinions were different according to the location of the store (Li & Fleyeh, 2018). Also on Twitter can be used to know the feelings about the COVID-19 of a country, the work done on the pandemic in Indonesia, investigated on the opinions of foreigners in the country, with 8740 tweets, finding that most messages were totally negative (Simanjuntak & Pramana, 2021). If research on COVID-19 is investigated, Post- COVID-19 opinion is also related, mostly because of vaccines, due to the fact that some people support them and others reject them, and which such information can help health institutions to mitigate fear or correct errors in vaccination (Alabid & Katheeth, 2021). Even bullying or school bullying can be detected, in Twitter 454 tweets were collected where it was possible to buy that 54.4% were tweets with bullying and only 35.4% were without bullying, although the accuracy is 60.5% (Khaira et al., 2020).

When talking about education has had a boom due to the pandemic, a study conducted in Ecuador for university students determined that 64% had a negative opinion, after applying 55 interviews, where problems such as stress and fatigue were shown (Ruben et al., 2020). Another constant evaluation is the change in education due to the transformation to online education, a research conducted in Jiangsu, with 1180 students showed that it was satisfactory with an index of 0.8205, approving the online education and the different platforms (J. Zhang, 2021). Also due to the pandemic it is important to know if by offering a program presents a positive message within the professional profile, which can be known by the mission, for example in Colombia the mechanical engineering program was analyzed in three universities in the department of Norte de Santander, where the positive index was higher than 60% (Castrillon et al., 2022).

This research analyzes the positive index of the Systems Engineering Program at the Universidad Francisco de Paula Santander Ocaña (UFPSO) in Colombia, through five factors: mission, vision, objectives, professional profile and occupational profile; which will allow knowing if through the information on the program's website, applicants can perceive a positive message that influences their selection, based on the frequency of positive, negative and neutral tokens.

2. METHODOLOGY

Five documents containing information about the Systems Engineering program are analyzed: Mission, Vision, Objectives, Professional Profile and Occupational Profile; for a total of 278

words that are converted into tokens (Table 1). Wolfram software is used for the Natural Language and sentiment analysis functions.

Table 1. Documents and number of words

Document	**Number of words**
Mission	74
Vision	47
Objectives	110
Professional Profile	68
Occupational Profile	98
Total tokens	**397**

Figure 1 shows the steps performed for the analysis of all documents, which begins by reviewing each document on the UFPSO program web page, then creating a vector with each document where the words are converted to tokens, and deleting insignificant tokens, converting all tokens to lowercase, measuring the final size of tokens, continuing with the calculation of the percentage of positive sentiment based on equation 1.

(1) $$\text{Tokens evaluate} = \text{Significant tokens} - \text{insignificant tokens}$$

(2) $$\text{Positive index} = \frac{\#\text{Positives}}{\#\text{negatives} + \#\text{Positives}}$$

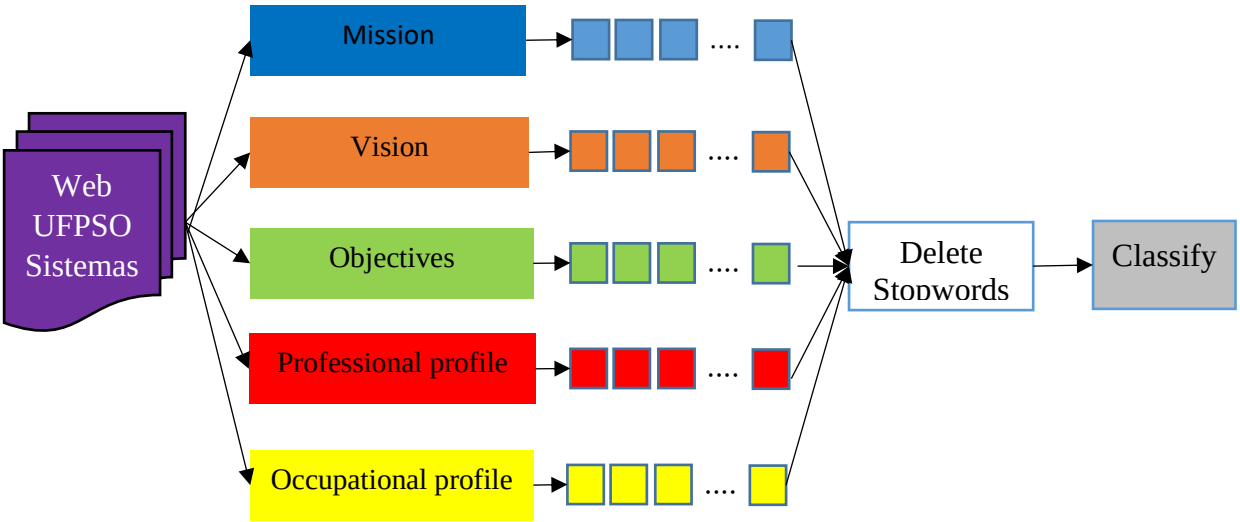

Figure 1. Sentiment analysis process.

The process to extract the information is captured from the program's website, Mission (UFPSO, 2020b) is as follows:

> "To train integral professionals in Engineering Computer Science Software Engineering and Information Technology Infrastructure (IT) with a solid humanistic formation that responds to the current problems and future needs of the region to face the technological challenges in accordance with the new information and knowledge society"

The Vision (UFPSO, 2020e) is as follows:

> "The Systems Engineering program will be recognized regionally as a leading program in the development of solutions in the area of Computer Science, Software Engineering and IT Infrastructure, through knowledge management and research, promoting humanistic, scientific and technological development, interpreting the globalized context in which it operates."

The objectives (UFPSO, 2020c) are as follows:

> "To develop in teachers and students a sense of participatory democracy, feelings of solidarity, awareness of social reality, spirit of justice, will for peace, continuous effort and teamwork, as ways to solve community problems.
>
> - To position social projection as an integral service that improves the quality of education and the quality of life, confronting theory and practice in real applications and with positive results for all.
> - Contribute to the improvement of personal competencies and to the improvement and updating of professional skills and knowledge.
> - To improve the deepening of human relations, the understanding of the regional environment, the proper interpretation of world events, mobilization, innovation, leadership and community participation."

The professional profile (UFPSO, 2020d) is as follows:

> "The UFPSO Systems Engineer is an integral professional trained under ethical principles, innovative, with an analytical-critical thinking capable of adapting and transforming the environment through applied research proposals; with an understanding of systems from the interaction between people, processes, and technology. The systems professional has the ability to conceive, design, implement and operate technological solutions that contribute to the achievement of the objectives of people and organizations."

The Occupational Profile (UFPSO, 2020a) is as follows:

> "At the end of his career the Systems Engineer graduated from UFPSO can perform within the following professional perspectives:
>
> - Advisor and consultant for the development of computer systems projects that contribute to Local, Regional and National economic growth.
> - As a Software Engineer creates solutions that are aligned with the objectives and requirements of the organization.

- Leads processes in the area of computing that provide technological solutions for its strengthening.
- Promotes the development of the region in a comprehensive way favoring the general interest of society.
- Generates innovative solutions through his research training to his work and training environment."

3. RESULTS

After applying Equation 1 to the 5 documents and extracting the final tokens to analyze, the results show that 397 tokens are reduced to 212 significant tokens, with 53.40% used (Table 2).

Table 2. Final tokens to be analyzed

Document	**Full tokens**	**Significant tokens**
Mission	74	28
Vision	47	27
Objectives	110	66
Professional Profile	68	38
Occupational Profile	98	53
Total tokens	**397**	**212**

For the Mission, 28 tokens were finally analyzed, where the index is positive with 59%, it can be seen in Figure 2 the word cloud, where Engineering and information are the most prominent in the message. Figure 3 also contains many neutral tokens which can affect the positive index of the message, with 13 positive, 9 negative and 6 neutral.

$$\text{P_mission} = \frac{13}{9 + 13} = 0.59$$

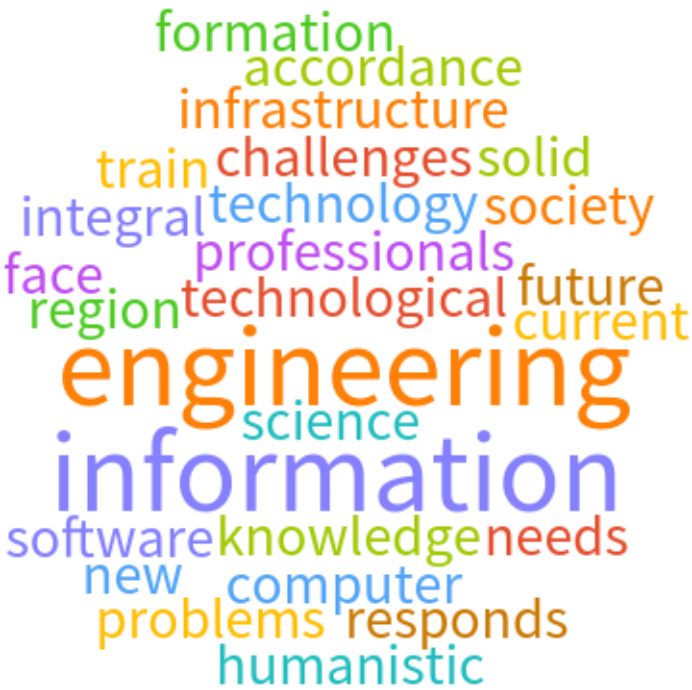

Figure 2. Mission word cloud.

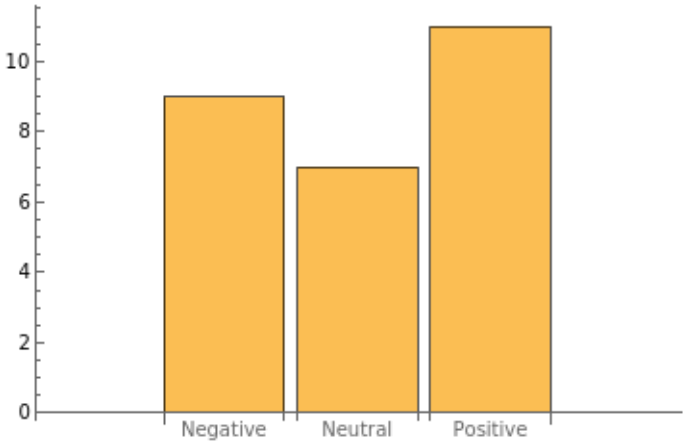

Figure 3. Mission bar chart

For the Vision, 57.44% of the tokens are analyzed for being the most significant, with a final positive index of 55%, with 11 positive, 9 negative and 7 neutral tokens; the word cloud shows a predominance of the words engineering, development and program (Figure 4). In the vision bart chart, the number of neutral words is still high, and the difference between positive and negative tokens is not significant (Figure 5).

$$\text{P_vision} = \frac{11}{9 + 11} = 0.55$$

Figure 4. Vision word cloud.

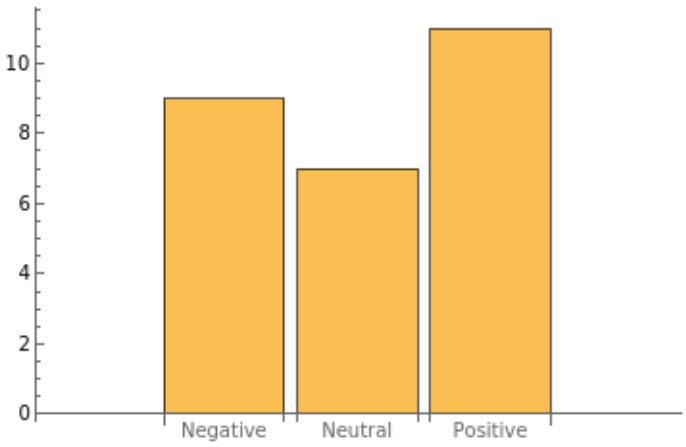

Figure 5. Vision bar chart

For the objectives, the total number of significant tokens was 66, with 60% of all tokens. A positive index of 64% was obtained, where the most predominant tokens were quality, improvement, community and social (Figure 6). There were 31 positive tokens, 17 negative, 14 neutral and 1 undetermined (Figure 7), where the message generated a more positive tendency than the program's mission and vision.

$$\text{P_objectives} = \frac{31}{17 + 31} = 0.64$$

Figure 6. Objectives word cloud.

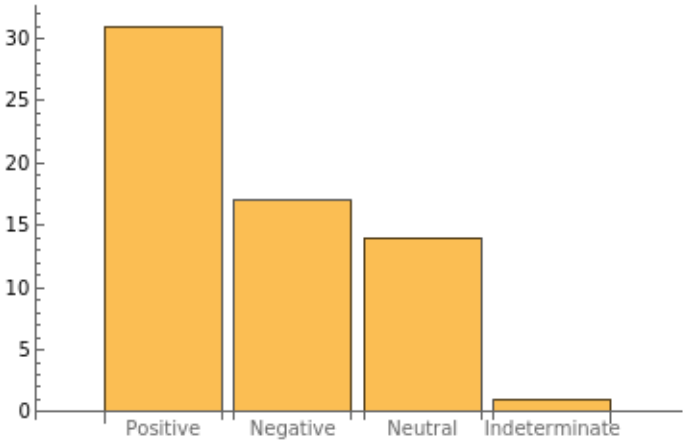

Figure 7. Objectives bar chart

The professional profile has 38 significant tokens and the occupational profile has 53, with a positive index of 61% and 65%, with more defined and clearly positive messages, where the items are defined according to the profile, where in the word cloud the tokens of professional, systems, people, training, solutions and development are widely spread (Figure 8). The professional profile contains 19 positive, 12 negative, 6 neutral and 1 indeterminate tokens, while the occupational profile has 28 positive, 15 negative and 10 neutral tokens (Figure 9).

$$\text{P_Professional} = \frac{19}{12 + 19} = 0.61$$

$$\text{P_Occupational} = \frac{28}{15 + 28} = 0.65$$

Figure 8. Word Cloud Profile - Professional (left) and Occupational (right)

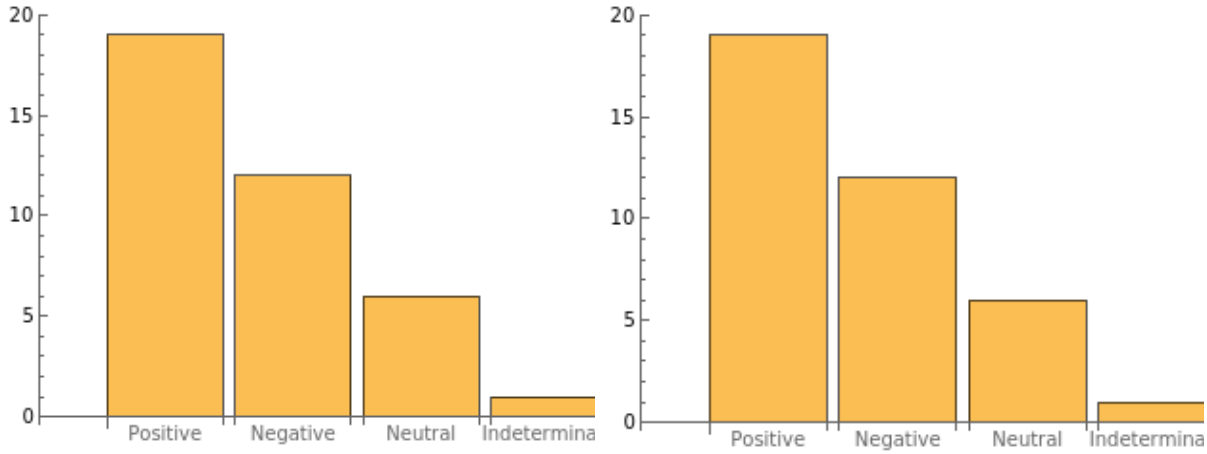

Figure 9. Bar chart Profile - Professional (left) and Occupational (right)

The average of the 5 documents has a positive index of Mission 59%, Vision 55%, Objectives 64%, Professional Profile 61% and Occupational Profile 65% (Figure 10); which shows a positive index in the 5 messages presented by the systems engineering program, however there is room for improvement to raise the ranks to values with more hits, which can call or send a positive message about the career to new applicants, it is described that the mission and vision present very modest percentages that should be valued, perhaps by containing fewer words should be analyzed so that the message does not get lost or become a neutral message without information.

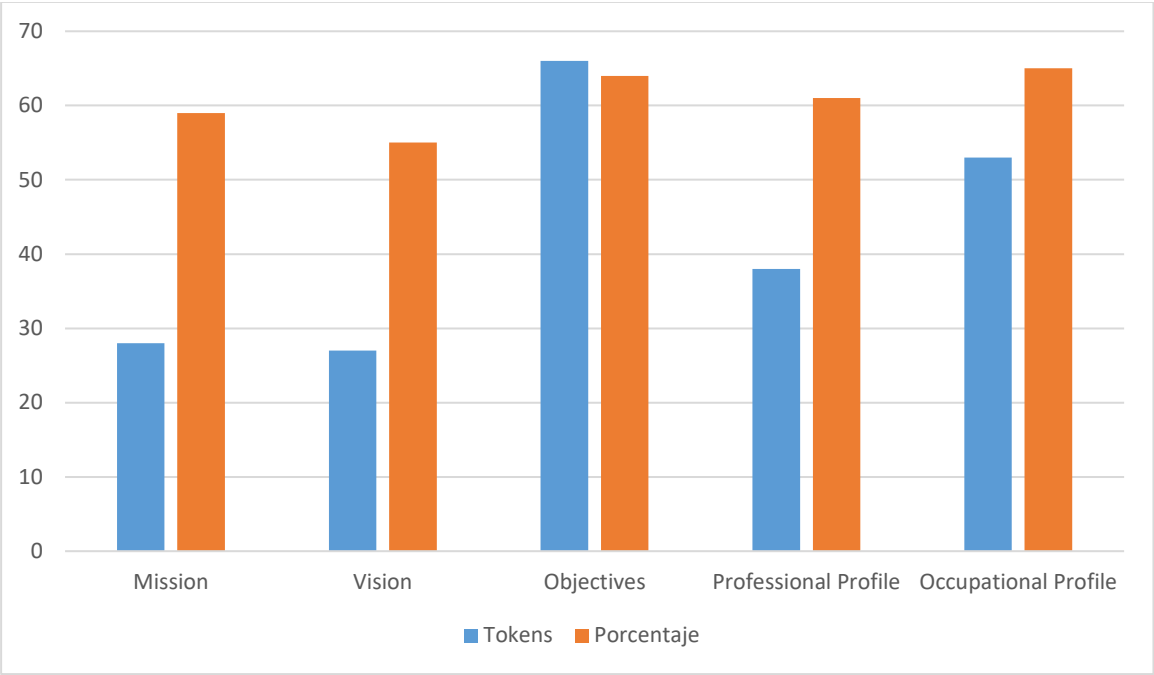

Figure 10. Percentage positive mentions

4. CONCLUSIONS

The semantic analysis of texts is a very important tool that allows to know the positive and negative index of text messages, and can be used in a wide variety of sectors and areas of industry, education, social networks and more. The analysis applied to the systems engineering program and the five factors described above show a positive index above 55%, and that the factors with the best weighting are the professional and occupational profile, however factors such as the mission and vision which are generally exposed first in all advertising guides show the lowest indexes. It is important to know that the five factors are found on the program's web page and today they are highly used due to the online enrollment to enter the university.

REFERENCES

Alabid, N. N., & Katheeth, Z. D. (2021). Sentiment analysis of Twitter posts related to the COVID-19 vaccines. Indonesian Journal of Electrical Engineering and Computer Science, 24(3), Art. 3. https://doi.org/10.11591/ijeecs.v24.i3.pp1727-1734

Alzboon, M. (2022). Semantic Text Analysis on Social Networks and Data Processing: Review and Future Directions. Information Sciences Letters, 11(5). https://digitalcommons.aaru.edu.jo/isl/vol11/iss5/6

Castrillon, S.-A. S., Camperos, J. A. G., & Jaramillo, H. Y. (2022). Sentiment Analysis Of The Educational Mission Of The Mechanical Engineering Programs In Norte De Santander. Journal of Language and Linguistic Studies, 18(4), Art. 4. https://jlls.org/index.php/jlls/article/view/5019

Cirqueira, D., Carmo, F., Cakir, G., Jacob Junior, A., Lobato, F., Bezbradica, M., & Helfert, M. (2020, noviembre 23). Explainable Sentiment Analysis Application for Social Media Crisis Management in Retail. https://doi.org/10.5220/0010215303190328

Khaira, U., Johanda, R., Utomo, P. E. P., & Suratno, T. (2020). Sentiment Analysis Of Cyberbullying On Twitter Using SentiStrength. Indonesian Journal of Artificial Intelligence and Data Mining, 3(1), Art. 1. https://doi.org/10.24014/ijaidm.v3i1.9145

Li, Y., & Fleyeh, H. (2018). Twitter Sentiment Analysis of New IKEA Stores Using Machine Learning. 2018 International Conference on Computer and Applications (ICCA), 4-11. https://doi.org/10.1109/COMAPP.2018.8460277

Lin, J.-K., Chien, T.-W., Yeh, Y.-T., Ho, S. Y.-C., & Chou, W. (2022). Using sentiment analysis to identify similarities and differences in research topics and medical subject headings (MeSH terms) between Medicine (Baltimore) and the Journal of the Formosan Medical Association (JFMA) in 2020: A bibliometric study. Medicine, 101(11), e29029. https://doi.org/10.1097/MD.0000000000029029

Medagoda, N., & Shanmuganathan, S. (2015, agosto 15). Keywords Based Temporal Sentiment Analysis. https://doi.org/10.1109/FSKD.2015.7382152

Ng, C., Lam, S., & Liu, K. (2022). Sentiment analysis on consumers' opinions – evaluating online retailers through analyzing sentiment for face masks during COVID-19 pandemic. Journal of Industrial and Production Engineering, 39(7), 535-551. https://doi.org/10.1080/21681015.2022.2070933

Rubén, P.-M., Patricio, B., Conde-González, M., & Peñalvo, G. (2020). Ecuadorian Higher Education in COVID-19: A Sentiment Analysis. ACM.

Setiawan, W. (2022, marzo 14). Text mining and semantics: A systematic mapping study Journal of the Brazilian Computer Society Full Text -. https://linisulsel.com/text-mining-and-semantics-a-systematic-mapping/

Simanjuntak, T. N., & Pramana, S. (2021). Sentiment Analysis on Overseas Tweets on the Impact of COVID-19 in Indonesia. Indonesian Journal of Statistics and Its Applications, 5(2), Art. 2. https://doi.org/10.29244/ijsa.v5i2p304-313

UFPSO. (2020a). Ocupacional. https://ufpso.edu.co/Ofertaufpso

UFPSO. (2020b). UFPSO - Mision. Mision. https://ufpso.edu.co/ofertaufpso/sistemas/Mision

UFPSO. (2020c). UFPSO - Objetivos. Objetivos. https://ufpso.edu.co/ofertaufpso/sistemas/Objetivos

UFPSO. (2020d). UFPSO - Pprofesional. Pprofesional. https://ufpso.edu.co/Ofertaufpso

UFPSO. (2020e). UFPSO - Vision. Vision. https://ufpso.edu.co/ofertaufpso/sistemas/Vision

Zhang, J. (2021). Research on Sentiment Analysis and Satisfaction Evaluation of Online Teaching in Universities During Epidemic Prevention. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.738776

Zhang, P., & Pan, Y. (2020). A Comparative Study of Keywords and Sentiments of Abstracts by Python Programs. Open Journal of Modern Linguistics, 10(6), Art. 6. https://doi.org/10.4236/ojml.2020.106044