

# Support System for Exchange Students for the Validation of Courses using Text Mining

Solange E. Barreda Muñoz<sup>1</sup>, Sebastian E. Quiroz Cervantes<sup>1</sup>, Jorge L. Martínez Muñoz<sup>1</sup> and Jose A. Sulla-Torres<sup>1</sup>

<sup>1</sup> Universidad Católica de Santa María, Urb. San José s/n Umacollo, Arequipa, 04000, Perú

## Abstract

Universities usually have a service to carry out student exchanges through the agreements established between the institutions. This service follows specific steps, including reviewing the syllabi, which generally takes a long time due to inconvenience. That is why an alternative solution is proposed with the use of text mining to compare the syllabus of each course and thus ensure an optimal and effective validation of courses. For this, the proposal has been based on the Rational Unified Process following all its phases and the R language for its implementation with a case study of a university in the city of Arequipa-Peru. The results showed that it was possible to find the most frequent and related words in the similarity of the syllabic documents. Therefore, it is concluded that the contribution provided by Text Mining helped improve the process of validating courses for student exchange between institutions.

## Keywords

Text mining, student exchange, syllabus comparison, assessment, text analysis.

## 1. Introduction

According to data from Peruvian universities, many undergraduate and postgraduate students exchange at a foreign university through an established agreement each year. This program has more than 30 destination countries. It has a series of requirements, among them validating the syllabus if they are going to study in such a way that they have the opportunity of an international experience [1].

The efficiency in the academic and administrative procedures of higher education marks the competitive advantage in quality aspects; in that sense, the treatment of the study plans (syllabus) in many cases is done manually, delaying many educational processes [2].

The problem lies when in the exchange process, the student goes to the destination university and has to choose the courses according to the name, which does not guarantee the similarity between them, an aspect that is considered ineffective since the name does not ensure that the content of one course is the same or very similar to another.

This makes the experience less attractive since, among other objectives, the student gets to learn more by living an experience abroad and not generate inconveniences with bureaucratic or cumbersome procedures.

The process of carrying out the exchange is often not automated; it requires certain administrative areas such as Cooperation and International Relations Offices, the Directorate of the Professional School, and the Parts Table, which means that the time it takes This process takes more than a week due to other internal processes that are considered as requirements to deliver the complete documents, for example, one of the requirements is a psychological study. On the other hand, for the universities, it is a situation that is expected to be feasible since they are the ones that must accept the exchange and communicate it both to the respective offices and the student. Likewise, the Cooperation and International Relations Offices must inform the School Director about the student's exchange before their trip.

---

*CITIE 2022: International Congress of Trends in Educational Innovation, November 08–10, 2022, Arequipa, Peru*

EMAIL: 72035128@ucsm.edu.pe (A. 1); 71569311@ucsm.edu.pe (A. 2); jmartinez@ucsm.edu.pe (A. 3); jsullato@ucsm.edu.pe (A. 4)

ORCID: 0000-0001-5353-184X (A. 1); 0000-0001-8743-4255 (A. 2); 0000-0003-0229-3508 (A. 3); 0000-0001-5129-430X (A. 4)



© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

On the other hand, text mining is increasingly helpful in the treatment of documentation in different fields and in the handling of information in documentary procedures. The text mining methodology can be used for educational tasks [3].

The proposal aims to help students compare syllables through text mining, which through the keyword frequency approach, supports a variety of operations in the Knowledge Discovery in Textual Databases (KDT), providing a base suitable for knowledge discovery and exploration for collections of unstructured text. This research also seeks the construction of the WebApp called SwapMe to contribute to the automation and improvement of the university exchange system.

## 2. Literature review

Among the different investigations on the subject, the article on text mining in education [4] was reviewed, which talks about the growth of online education, which produces exciting challenges on how to extract text data to find valuable knowledge for those interested in education. This paper helps us to see text mining is used and how it will be applied in the educational environment. The main objective is to answer three research questions: What are the most used text-mining techniques in educational settings? What are the most used educational resources? Moreover, what are the main applications or educational objectives?

In the paper by Feldman, Dagan, and Hirsh [5] on Text Mining Using Keyword Distributions, they describe the KDT system for knowledge discovery in text, in which keywords tag documents, and knowledge discovery is This is done by analyzing the frequencies of simultaneous occurrence of the various keywords that tag the documents.

Kadampur and Riyae [6] present a web-enabled, AI-powered, personalized service application for educators to automatically configure a quiz on the selected topic or syllabus. The AI component of the app works on text mining and content classification. It also helps suggest questions about the content of the supplied text.

In the article by Ito et al. [7], they propose a project-based learning educational system that uses artificial intelligence instead of direct instruction from teachers. For this, they used the e-Syllabus that the teachers have used as a communication tool with the students. Some have been using the e-Syllabus by carrying out flipped and active learning [8]. They consider text mining vital as it transforms unstructured text into structured words and extracts meaningful patterns. You can explore and discover new meanings in the text data.

In the work of De Aires, Corría and Godinho [9] explored how student engagement can be promoted through transmedia using a set of activities within the Moodle learning management system for a syllabus topic on innovation for an entire semester. To perform the data analysis, we followed a mixed method approach between descriptive statistics, data mining analysis based on the Orange open-source software, and a final questionnaire.

Kaibassova et al. [10] review the use of intellectual data analysis methods to form educational programs in the context of determining the sequence of studying disciplines in the direction for consideration. The article succinctly describes the developed software application that allows extracting information from text documents, processing, analyzing and visualizing data. The proposed model performs the grouping of text documents taking into account the weighting coefficient of the individual words of the corpus.

According to Kawintiranon et al. [11] indicate that curriculum analysis is attracting widespread interest in the educational field. Point out two main approaches: (i) human-based and (ii) text-based assessments. They feature an automated text-based curriculum analysis that directly assesses complete course materials. The approach employs a well-known text mining technique that extracts keywords using TF-IDF. The analysis is based on the keywords in the course materials that match the keywords in the online documents, which is like the domain expert. Another similar study was carried out in [12] and also for curriculum validation [13].

Yasukawa, Yokouchi, and Yamazaki [14] investigated the searchability of a collection of curricula and compared methods for word suggestions using deep learning approaches and large text corpora. In the experiment, they used a bibliographic database from university libraries in Japan. The results indicated that a wide range of vocabulary is advantageous in improving the searchability of syllabuses.

Ajmal and Tak [15] present an efficient text-mining method that focuses on extracting and updating unknown words to improve data classification and POS tags. The System's main feature is finding such unknown foreign words and updating them to the appropriate words, which depends on the information available through the dictionaries. The proposed methods can also help improve the accuracy of extracting frequent patterns and association rules from unstructured (textual) data.

### 3. Methodology

The Project is based on the Rational Unified Process (RUP) methodology [16]. According to the Project's characteristics, the participants' roles, the activities to be carried out, and the artifacts (deliverables) were selected. The four phases will be fulfilled, observed in Figure 1, which marks the methodology, consisting of the phases of initiation and elaboration: one iteration, three iterations of the construction phase, and two iterations of the transition phase.



Figure 1: Phases of the RUP Method

#### 3.1. Phase 1: Design

In this initiation phase, the workflows are shown as the main requirements (see Table 1) to define and agree on the Project's scope with the interested parties and identify the possible risks for the Project, shown in Table 2.

Table 1  
Main Requirements

Request	Description
Authentication	It will allow the user to enter the System through an authentication method
Visualization of Agreements	The application will show the agreements that the University has
Visualization of Universities	The System will allow you to see the universities in which the exchange can be carried out
University selection	You can select the University where you want to do the exchange
Racing information display	After selecting the University, the career will be selected, and the study plan can be seen
Course Comparison	Once the destination university has been selected, the course comparison process can be carried out

Text Mining	These techniques will be used for the development of the comparison process
Comments and ratings	Universities will have a comments section where students will talk about their experiences and give an assessment of their exchange
Change of language	The System will be able to change the language depending on the user's preference

**Table 2**  
Risks external to the Project

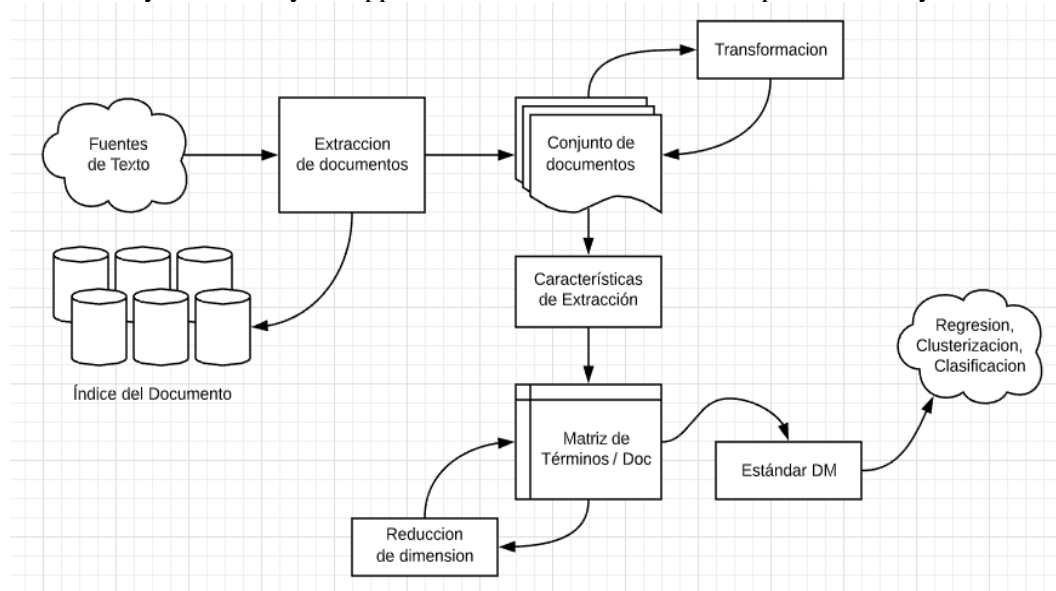
Risk ID	Description	Risk
R01	Resistance to change	Scope not accepted by the client
R02	Information obtained out of date	Quality failures in the System
R03	The end user does not understand the operation of the System	Scope not accepted by the client
R04	Lack of professional career curricula	Do not possess the syllables; without them, the System will not be able to function

### 3.2. Phase 2: Elaboration

In the elaboration phase, we focus on the design and analysis aspects to verify that the Project is viable and to know the technologies that will be used and, as the main result, the obtaining of a stable architecture.

#### 3.2.1. Technology

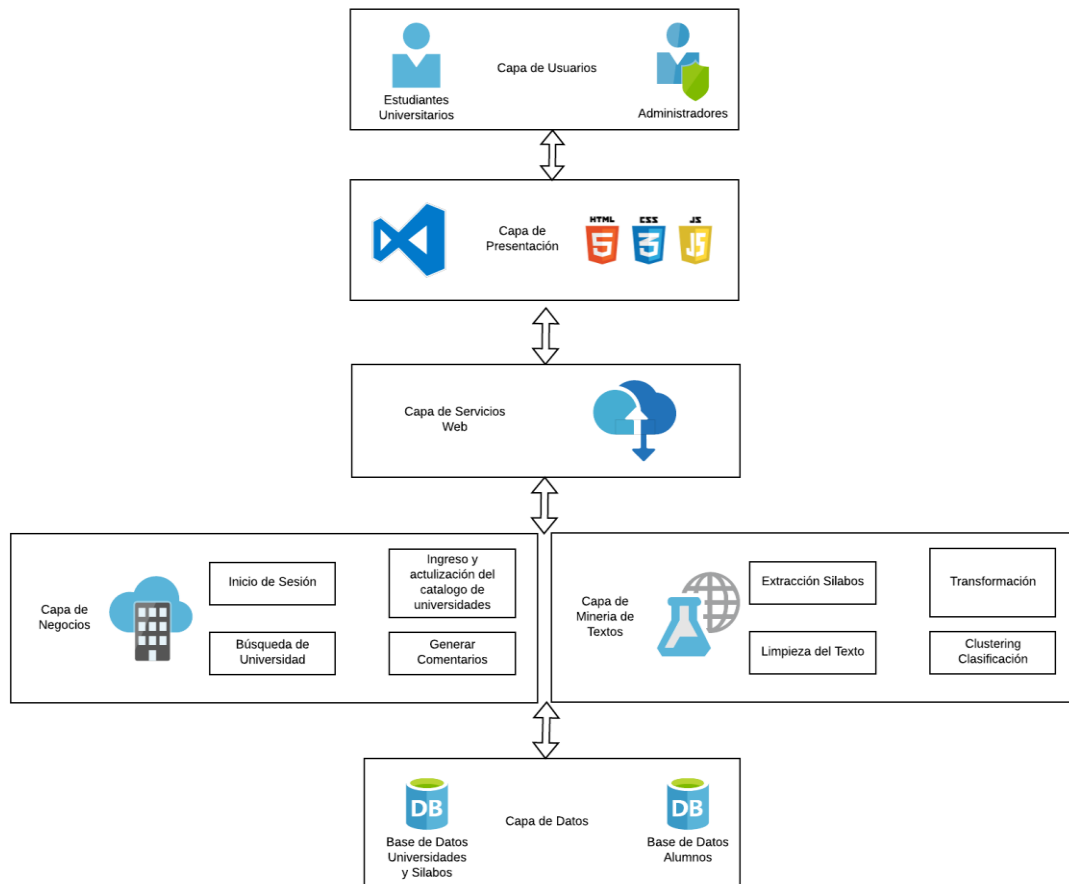
The most important technology is text mining as a life cycle [17] for elaborating the proposal. Figure 2 shows the life cycle of how your application will be carried out compared to the syllabus.



**Figure 2:** Text Mining Lifecycle

#### 3.2.2. Architecture

The description of the Logical Architecture is shown in Figure 3; it is composed of 5 layers where we have the most important aspects that we will describe below.



**Figure 3:** Logical architecture of the proposal.

1. **User Layer:** In this layer, you can see the actors that interact with the System; in the architecture, we can identify them as students and administrators, both can access the System, but it has different functionalities available.
2. **Presentation Layer:** In this layer, the technologies that are going to be used for the development of the application are presented, which in this case were already defined as the classic web technologies that are HTML, CSS, and JAVASCRIPT, and that will be implemented with the use of VSCode, which is our selected text editor and which facilitates the development of code thanks to its intelligent autocompletion and extensions such as Live Server that allows us to see the visualization of all the changes that we are making.
3. **Web Service Layer:** In this layer are the various web services that will provide us with a secure way to access the System and the databases.
4. **Business Layer:** In this layer are all the processes that users can use, from entering the System and validating their credentials to requesting the comparison of syllables between different universities for students and, in the case of administrators, adding new universities, racing, and other options.
5. **Text Mining Layer:** In this layer, the text mining process is carried out in order to verify the compatibility between careers, where the syllables to be compared are first extracted; after having it, they are cleaned, that is, we eliminate words such as articles, punctuation marks, and connectors, in order to obtain the keywords and then be able to carry out the clustering and classification processes [18] in order to obtain a percentage of similarity and ensure that the same topics are covered in both universities and ensure that the student can carry out the exchange with more security.

- Data Layer: In this layer, the Database will be used for the correct functioning of the application. We have the University's Database and the other databases of the other universities with which it can be done exchanges.

### 3.3. Phase 3: Construction

In this phase, the priority is to reach the operational capacity of the product incrementally with iterations; we are dedicated to mitigating all the risks found in the development phase and implementing all the characteristics and requirements previously described to obtain a version of the WebApp called SwapMe. Figure 4 shows the Class Diagram developed.

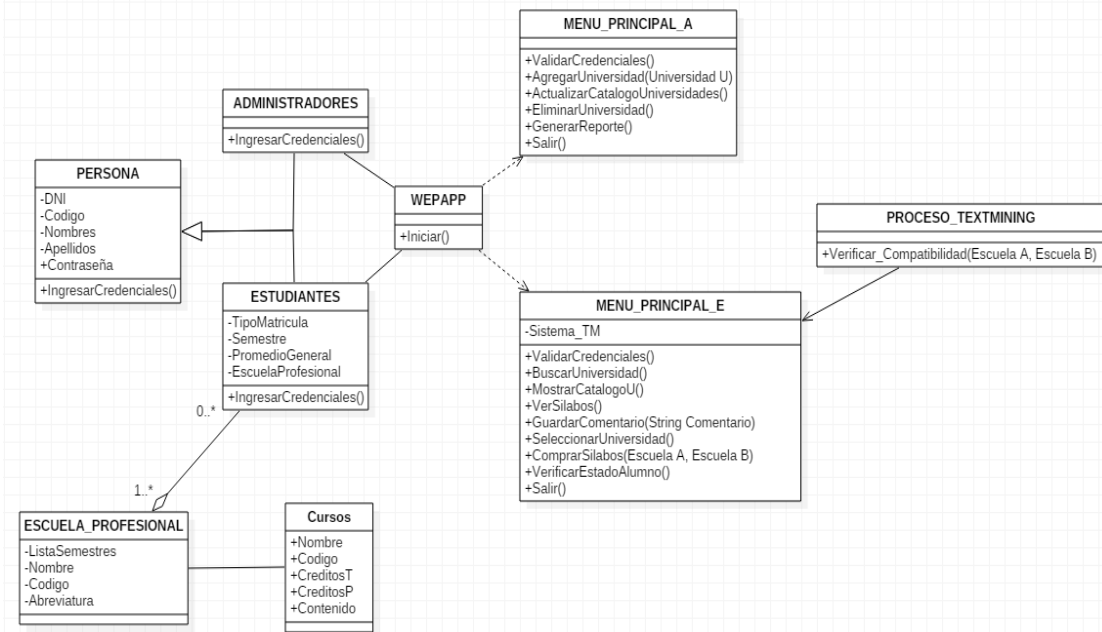


Figure 4: Class Diagram.

The materials used were the R Studio tool [19] with its R language, which is a free and open-source multiplatform programming environment. R was specifically designed for statistical analysis, which makes it well-suited for data science applications. The tools it provides in R make it easy to perform robust text analysis. Shiny is a popular R package that makes it easy to create interactive web applications directly from R. You can host stand-alone applications on a web page, embed them in R Markdown documents, or create dashboards. The Database was managed using the SQL Server 2017 tool to administer student data, teachers, courses, syllables of each course, and University agreements.

#### 3.3.1. Syllabus Text Mining

In order to extract information about the courses from the documents provided by the universities, we will use text mining. To carry out the comparison process, we have used the R programming language, which has a statistical analysis approach focused on business intelligence that has algorithms for data mining development.

Text analysis, in particular, has stabilized mainly in R. There is a vast collection of libraries devoted to text processing and analysis, from low-level string operations to advanced-level text modeling techniques example: the adjustment of allocation models.

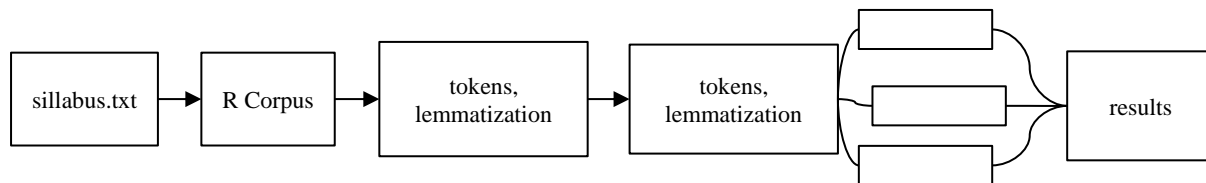
One of the main advantages of performing text analysis in R is that it is usually accessible and relatively not so complicated to combine packages or libraries, where the complexity of analyzing is seen in the interpretation of each of the calculations that are performed when handling the imported data. This challenge plays an increasingly important role for developers in terms of cooperation and coordination. Recent research among the R text analytics developer community is designed to promote

interoperability to increase implementation flexibility on this topic and thus result in learning basic implementation concepts by parsing texts and providing access to a wide range of advanced functionality for this challenge. Table 3 shows the structure of the operations in the R packages to be used.

**Table 3**  
Structure of package operations in R

Operation	Library	Options
Data Preparation	readtext	
Import of Texts	stringi	txt, readxl, pdftools
String Operations	quanteda	stringr
Preprocessing	quanteda	stringi,tokenizers,snowballC,tm
DTM	quanteda	tm, tidytext, matrix
Filtration and weighing	quanteda	tm, tidytext, matrix
Analysis	quanteda	
Dictionary	quanteda	tm, tidytext
Supervised Machine learning	quanteda	RTextTools,kerasR
Unsupervised Machine Learning	topicmodels	quanteda, stm, text2vec

In the table above, you can see the operations in a sequence of how the texts should be analyzed. For a better understanding, Figure 5 shows the diagram of the synthesized steps of the analysis of the syllables:



**Figure 5:** Stages of syllabus analysis.

First, we chose two syllabi from different universities with similar names to analyze their content. Before comparing the text of both syllables, it is necessary to carry out the first step of text mining: data preprocessing.

We need to install a few packages in RStudio, which is the tool we will use to develop with the R language.

After installing these packages, we define the folder where we will get the text from and import the packages to the created Project; each package has different functionality, such as:

- tm: Specific for text mining.
- wordcloud: To graph the word cloud that we will see later.
- dplyr: With functions to manipulate and transform the data, it allows you to change or delete the operator so that the text is more readable.
- readr: Allows you to read and write documents.
- cluster: To perform group analysis.

Once the file is read, we define from which line block the document will be read. We are left with elements equal to the number of lines read in the document and the selected range. We combine the column number with what each column contains. It is also necessary to clean the data by converting everything to lowercase, eliminating empty words, among others, eliminating the numbers because, in this case, they are not necessary, and finally, the blank spaces.

We convert everything to plain text and draw the word cloud [20] shown in Figure 6. It is necessary to eliminate specific words that are repeated so that the words shown at the end are essential and have more meaning. For example, the word practice or phase, in this case, is not applicable, so we eliminate it and redraw the word cloud.



**Figure 6:** Generated word cloud.

In this field of text mining, we usually start with a set of records that, in the R language, are called highly heterogeneous input documents. For this reason, the first step to execute is to import it into a computational environment.

It is basically what we will use R for since this tool with the tm package is a standard for statistical analysis of text in R.

Once the files have been imported, which in this case have been converted to .txt, the first cleaning of the imported data is carried out to avoid noise when analyzing it; that is, remove punctuation marks, integers and decimals, blank spaces, and, finally, as a cleaning standard, the texts are converted to lowercase.

Now with the first cleaning to be able to evaluate these texts, different structures will be needed:

The first structure is considered data.frame, a data frame (table) of two-dimensional matrix type in which each column has a variable's values and each row has a set of values in the respective columns.

**Table 4**

Structure of the data frame

N°	id	text
1	ucsm.txt	(Syllabus content 1)
2	alas.txt	(Syllabus content 2)

The second is considered similar to that of a database that will contain and help to manipulate the documents in a general way, shown in Table 5. This structure is called corpus [10], found in both the tm and quanteda libraries. For our Project, we will use both since that way; we will properly clean the imported data, as well as make it easier for us to handle it.

**Table 5**

Corpus structure

library	structure
tm	id - text (secondcleaning: to_Lower)
quanteda	Corpus consisting of 2 documents and 2 docvars

Once the first part of the general structure of the documents has been built, a total of documents with the characteristic of being high-dimensional is obtained; that is, large.

Since we already have the corpus, the next step is its processing, which means applying the methods to finish cleaning and structuring the input, thus also identifying the characteristics in a simplified set that can be analyzed later.



Among the applications to use in the text are: ID1 => `tm_map(corpusCompleto,stripWhitespace)`, y ID2 => `tm_map(corpusCompleto,removeWords,stopwords("spanish"))`, where the white spaces will be eliminated (ID1) and an important part the elimination of empty words that are considered in the Spanish language as prepositions.

From what has already been captured and obtained (corpus), one of the essential points when delving deeper into data mining is the creation of the document-terms matrix, which contains rows that correspond to the imported documents and the columns to their terms, which is shown in Figure 7.

dtm	list [2 x 467] (S3: DocumentTerm	List of length 6
i	integer [440]	1 1 1 1 1 1 ...
j	integer [440]	1 3 4 5 6 8 ...
v	double [440]	0.000881 0.000881 0.002.
nrow	integer [1]	2
ncol	integer [1]	467
dimnames	list [2]	List of length 2

Figure 7: document-term matrix

### 3.4. Phase 4: Transition

The final stage of RUP, where the product obtained after all the iterations carried out, is placed in the hands of the end users to verify that our product obtained meets all the standard requirements and satisfies the users correctly. The main deliverables are the operational prototype of the application and all the necessary documentation.

The WebApp software was made using the R language using the Shiny library, which allows the creation of web pages using HTML, CSS, and JavaScript. Figure 8 shows the main window of the built WebApp.

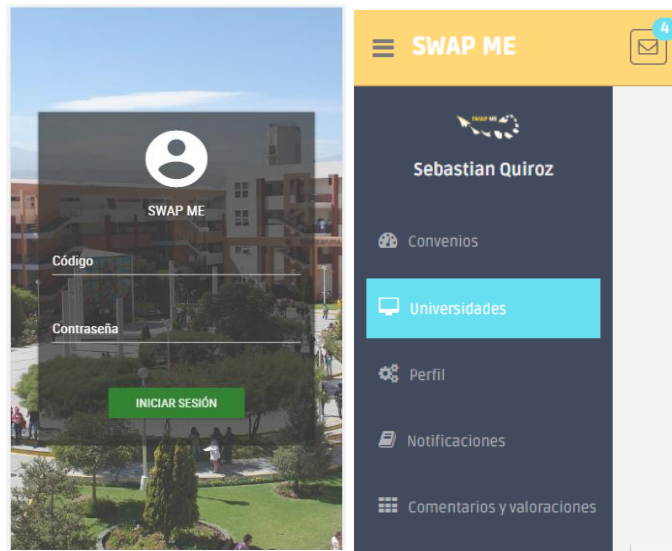


Figure 8: WebApp main window

Figure 9 shows the window of available universities and agreements.

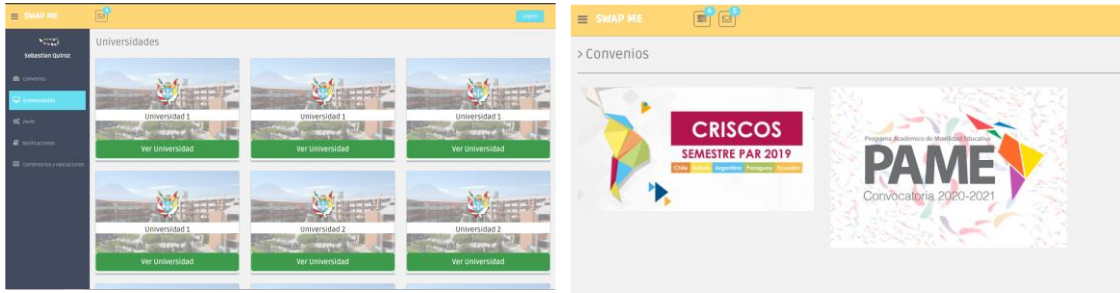


Figure 9: Window of available universities and agreements.

## 4. Results

In this section, the evaluation results of the built application were obtained.

The distance between the syllabic documents was found to see their similarity according to the following equation 1.

$$similarity(doc_1, doc_2) = \cos(\theta) = \frac{doc_1, doc_2}{|doc_1||doc_2|} \quad (1)$$

```
> d1_d2_cos_sim
2 x 2 sparse Matrix of class "dsCMatrix"
      text1 text2
text1  1     0
text2  0     1
```

According to the Euclidean distance, it shows the similarity percentage: text1 text2 is 30.09983. The word frequencies obtained in the tests are shown in Figure 10.

	palabra	frec
<b>móviles</b>	móviles	25
<b>desarrollo</b>	desarrollo	16
<b>aplicaciones</b>	aplicaciones	15
<b>android</b>	android	11
<b>proyecto</b>	proyecto	11
<b>trabajos</b>	trabajos	11
<b>conceptos</b>	conceptos	10
<b>grupo</b>	grupo	10
<b>responsabilidad</b>	responsabilidad	10
<b>temas</b>	temas	10
<b>presentación</b>	presentación	9
<b>saber</b>	saber	9

Figure 10: Word frequency

Another aspect of interest when analyzing these texts is seeing the relationship-specific keywords with others. To do this, we are going to see next to which words the words appear: "intelligence," "learning," "diffuse," "deep," and "learning." Figure 11 shows the relationships between the words with similarity percentages.

```

$inteligencia
  artificial      big      computacional      data      desarrolladores
    0.99          0.95          0.95          0.95          0.95
  jaén          mathivet      implementación      desarrollo      encargado
    0.95          0.95          0.80          0.60          0.60
  universitaria      descripción
    0.30          0.26

$aprendizaje
  supervisado      automático      distancia      inductivo      inferencia
    0.83          0.82          0.71          0.71          0.71
  machine          medidas      modalidad      lógica      introducción
    0.71          0.71          0.71          0.49          0.47
  procedimental      aplicación      automotiva      clasificación      clustering
    0.47          0.45          0.45          0.45          0.45
  conjunto          evalúa      heurística      metaheurística      regresión
    0.45          0.45          0.45          0.45          0.45
  competencia      difusos      unidades      iii
    0.28          0.28          0.28          0.27

$difusa
  lógica      difusos      aplicación      automotiva      clasificación
    0.93          0.74          0.70          0.70          0.70
  clustering      conjunto      evalúa      heurística      metaheurística
    0.70          0.70          0.70          0.70          0.70
  regresión      automático      supervisado      actitudinal      aprender
    0.70          0.59          0.47          0.36          0.33
  autónomo      compromiso      computadoras      control      cuantificación
    0.33          0.33          0.33          0.33          0.33
  difuso      entiende      estudiantil      evidencia      función
    0.33          0.33          0.33          0.33          0.33
[ reached getOption("max.print") -- omitted 11 entries ]

```

**Figure 11:** Relationships between words with similarity percentage

## 5. Conclusions

Text mining is beneficial for any organization as it can save money and solve problems that can be used for proper decision-making.

In the process, the elimination of specific words that are repeated is an essential aspect since it selects words to show the most significant meaning, which was obtained according to the obtained frequency tables.

The relationship of keywords with others is another essential aspect since it allows knowing the similarity of keywords that appear and thus forming a basis for comparison through the distances used. The developed AppWeb, allowed us to have a better interaction in usability with the users in a personalized way through the interfaces. It is a very friendly and operative application for the student since we can use it both on the web and as an application.

Likewise, it can be concluded that this application would help institutions to be more efficient in the student exchange process since it is a more practical way to apply to the universities with which a university is affiliated. On the other hand, the most important thing is to make the agreements; it makes it easier for students to verify and validate if it is convenient to exchange to the chosen University.

## 6. References

1. ¿Por qué participar en un programa de intercambio? - Portal de Internacionalización | PUCP, <https://internacionalizacion.pucp.edu.pe/intercambio-estudiantil-pucp/por-que-participar-en-un-programa-de-intercambio/>, last accessed 2022/10/13.
2. Meza-Luque, A., Del Carpio, A.F., Paredes, K.R., Sulla-Torres, J.: Architectural proposal for a syllabus management system using the ISO/IEC/IEEE 42010. *Int. J. Adv. Comput. Sci. Appl.* (2020). <https://doi.org/10.14569/IJACSA.2020.0110640>.
3. Orellana, G., Orellana, M., Saquicela, V., Baculima, F., Piedra, N.: A text mining methodology to discover syllabi similarities among higher education institutions. In: *Proceedings - 3rd International Conference on Information Systems and Computer Science, INCISCOS 2018* (2018). <https://doi.org/10.1109/INCISCOS.2018.00045>.
4. Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., Romero, C.: Text mining in education, (2019). <https://doi.org/10.1002/widm.1332>.

5. Feldman, R., Dagan, I., Hirsh, H.: Mining text using keyword distributions, (1998). <https://doi.org/10.1023/A:1008623632443>.
6. Kadampur, M.A., Riyaae, S. AI: QPSetter: An Artificial Intelligence-Based Web Enabled, Personalized Service Application for Educators. In: Lecture Notes in Networks and Systems (2022). [https://doi.org/10.1007/978-3-030-82193-7\\_51](https://doi.org/10.1007/978-3-030-82193-7_51).
7. Ito, T., Tanaka, M.S., Shin, M., Miyazaki, K.: THE ONLINE PBL (PROJECT-BASED LEARNING) EDUCATION SYSTEM USING AI (ARTIFICIAL INTELLIGENCE). In: Proceedings of the 23rd International Conference on Engineering and Product Design Education, E and PDE 2021 (2021). <https://doi.org/10.35199/epde.2021.19>.
8. Ito, T., Ishii, K., Nishi, M., Shin, M., Miyazaki, K.: Comparison of the effects of the integrated learning environments between the social science and the mathematics. In: SEFI 47th Annual Conference: Varietas Delectat... Complexity is the New Normality, Proceedings (2020).
9. De Aires Angelino, F.J., Loureiro, S.M.C., Bilro, R.G.: Analysing students' engagement in higher education through transmedia and learning management systems: A text mining approach. *Int. J. Innov. Learn.* 30, (2021). <https://doi.org/10.1504/IJIL.2021.118875>.
10. Kaibassova, D., La, L., Smagulova, A., Lisitsyna, L., Shikov, A., Nurtay, M.: Methods and algorithms of analyzing syllabuses for educational programs forming intellectual System. *J. Theor. Appl. Inf. Technol.* 98, (2020).
11. Kawintiranon, K., Vateekul, P., Suchato, A., Punyabukkana, P.: Understanding knowledge areas in curriculum through text mining from course materials. In: Proceedings of 2016 IEEE International Conference on Teaching, Assessment and Learning for Engineering, TALE 2016 (2017). <https://doi.org/10.1109/TALE.2016.7851788>.
12. Föll, P., Thiesse, F.: Exploring Information Systems Curricula: A Text Mining Approach. *Bus. Inf. Syst. Eng.* 63, (2021). <https://doi.org/10.1007/s12599-021-00702-2>.
13. West, J.: Validating curriculum development using text mining. *Curric. J.* 28, (2017). <https://doi.org/10.1080/09585176.2016.1261719>.
14. Yasukawa, M., Yokouchi, H., Yamazaki, K.: Syllabus Mining for Faculty Development in Science and Engineering Courses. In: Proceedings - 2019 8th International Congress on Advanced Applied Informatics, IIAI-AAI 2019 (2019). <https://doi.org/10.1109/IIAI-AAI.2019.00074>.
15. Khan, I.A., Choi, J.T.: Lexicon-corpus Based Korean Unknown Foreign Word Extraction and Updating Using Syllable Identification. In: *Procedia Engineering* (2016). <https://doi.org/10.1016/j.proeng.2016.07.445>.
16. Jones, C.: 50 Rational Unified Process (RUP). In: *Software Methodologies A Quantitative Guide* (2017). <https://doi.org/10.1201/9781315314488-51>.
17. Castellano, M., Mastronardi, G., Aprile, A.: A web text mining flexible architecture. *World Acad. J.* (2007).
18. Nisa, R., Qamar, U.: A text mining based approach for web service classification. *Inf. Syst. E-bus. Manag.* 13, (2015). <https://doi.org/10.1007/s10257-014-0252-5>.
19. Martin, G.: R Studio. In: *An Introduction to Programming with R* (2021). [https://doi.org/10.1007/978-3-030-69664-1\\_1](https://doi.org/10.1007/978-3-030-69664-1_1).
20. Bao, C., Wang, Y.: A Survey of Word Cloud Visualization, (2021). <https://doi.org/10.3724/SP.J.1089.2021.18811>.