EasyChair Preprint

№ 3938

Analysis of Similarity in Written Works of ESCOM Students with the Turintin Technological Tool.

Monserrat Gabriela Pérez Vera, Sandra Mercedes Pérez Vera and Virginia Sánchez Cruz

July 23, 2020

Analysis of similarity in written works of ESCOM students with the Turintin technological tool.

Monserrat Gabriela Pérez Vera[1] Sandra Mercedes Pérez Vera[2], Virginia Sánchez Cruz[3]

[1] Integral and Institutional Training Department, Escuela Superior de Cómputo-IPN Av. Juan de Dios Bátiz s/n almost corner Miguel Othón de Mendizábal, Professional Unit "Adolfo López Mateos". Suburb Lindavista, C.P., del. Gustavo A. Madero, México City.
[1]mperezve@ipn.mx

[2] Integral and Institutional Training Department, Escuela Superior de Cómputo-IPN Av. Juan de Dios Bátiz s/n almost corner Miguel Othón de Mendizábal, Professional Unit "Adolfo López Mateos". Suburb Lindavista, C.P., del. Gustavo A. Madero, México City.
[2]sperezv@ipn.mx

[3] Department of Computer Systems, Escuela Superior de Cómputo-IPN Av. Juan de Dios Bátiz s/n almost corner Miguel Othón de Mendizábal, Professional Unit "Adolfo López Mateos". Suburb Lindavista, C.P., del. Gustavo A. Madero, México City.
[3]riomudo@yahoo.com.mx

Abstract. This work exposes the teaching and learning process of the 29 students enrolled in the Oral and Written Communication (COE) subject during the period 19-20 / 2, which is taught in the Computer Systems Engineering (ISC) degree.), this study program is offered at the Superior Computer School (ESCOM) of the National Polytechnic Institute (IPN). The students were asked to upload an activity to the Turnitin technological tool, which was the advancement of their research protocol, with the intention of identifying the similarity or plagiarism of their information, with the option of working on the tool until decrease the percentage of similarity of their work, obtaining as a result of similarity lower percentages in the final works of the students.

Key words: ESCOM students, similarity, plagiarism, Turnitin tool, written work**.**

1 Introduction

The present work exposes the teaching and learning process of the 29 students enrolled in the subject of Oral and Written Communication (COE) during the period 19-20 / 1, which is taught in the Computer Systems Engineering (ISC) degree.), this study program is offered at the Superior Computer School (ESCOM) of the National Polytechnic Institute (IPN); This subject belongs to the first semester of the degree, it should be noted that a large percentage of the students have been trained in the area of physical mathematics, where the subjects taken have been low where the training of oral communication skills is strengthened and written.

Students were asked to upload to Turnitin an activity to advance their research protocol, with the intention of identifying the similarity or plagiarism of their information, with the option of working on the tool to decrease the percentage of similarity of their job.

Turnitin is an application that helps you in detecting plagiarism, by comparing academic and professional works with a database, which includes (Recursos Digitales, s.f.).

The percentage of similarity of the works was decreased in a dizzying way, since all the students had the possibility of visualizing the percentage of similarity, making this a challenge for them and to attend to ethics as a student.

2 Background

Turnitin offers services to mitigate the risk of academic and professional plagiarism, as well as tools to support the teaching and learning process (Turnitin, s.f.).

According to Turnitin (2020): It is an internet plagiarism prevention service created by iParadigms, LLC, first launched in 1997. Typically, universities and institutes purchase licenses to submit essays to the Turnitin website, which check the document for non-original content.

In other words, Turnitin is an application that helps you detect plagiarism, by comparing academic and professional work with a database, which includes (Recursos Digitales, s.f.):

•Websites.
•Digital books.
• Press articles.
•Journals.
• Online publications.

It is relevant to note that since 1998 a solution has been created by four PhD students from the University of California at Berkeley, as a peer correction application (Turnitin, s.f.).

In 2000 Turnitin.com was launched and a plagiarism prevention service (Turnitin, s.f.) was introduced.

In 2002 an online qualification product was launched, after this, the Turnitin database will have one million jobs (Turnitin, s.f.).

In 2007 Turnitin partners with CrossRef to develop the largest academic database in the world, after which there will be 16 million registered users of Turnitin. In addition to that, the Turnitin database will grow to one hundred million deliveries (Turnitin, s.f.).

In 2010 Turnitin combines Originality Checking, Online Qualification and Peer Correction products in one service. Subsequently, Turnitin would purchase nLearning and establish offices in Newcastle, England (Turnitin, s.f.).

In 2014 Turnitin buys Ephorus and Lightside Labs, and establishes offices in Utrecht, Pittsburgh, and Austin. After that, 26 million students and educators would use Turnitin globally. In addition, the Turnitin database would grow to 500 million deliveries (Turnitin, s.f.).

In 2015 Turnitin exceeds 300 employees and would launch the Scoring Engine to instantly grade student writing (Turnitin, s.f.).

In 2016 Turnitin launched Revision Assistant to support student writing and Feedback Studio, a redesign of its core product (Turnitin, s.f.).

According to Turnitin (s.f.), Turnitin identifies unoriginal content through its similarity detection tool, aiding in the comparison of produced content the only database that includes the internet, student papers, and indexed scientific articles. In addition to assisting in the protection of intellectual property, it provides student tools for peer review and evaluation.

To meet its goals, Turnitin compares jobs submitted across multiple databases using a proprietary algorithm. Among the databases used are their own databases, as well as private academic databases (Turnitin, 2020).

According to Abad & Sevillano (2018) Turnitin compares the documents against:

- The material available on the open web, such as Wikipedia, El Rincón del Vago, etc. (Abad & Sevillano, 2018).
- Scientific academic documents hosted on platforms such as Elsevier, Science, Epsco, Emeral or Proquest (Abad & Sevillano, 2018).
- The works of the universities that are subscribed to this tool (Abad & Sevillano, 2018).

The operation of Turnitin according to Abad & Sevillano (2018) is as follows:

1. Users upload the documents they want to collate (Abad & Sevillano, 2018).
2. The tool searches for the text strings and compares them with the documents in its databases (Abad & Sevillano, 2018).
3. The user receives as a result the document they have compared and, in case of finding matches, the parts that match are indicated with a different color (Abad & Sevillano, 2018).
4. The final percentage is the percentage of all the text that appears in another document (Abad & Sevillano, 2018).

To know if there is plagiarism or not, according to Abad & Sevillano (2018) depends on the field of study:

- A 20% coincidence in fields like the right one is normal.
- A 3% coincidence in fields such as natural sciences may indicate possible copying.

However, it depends not only on the field of study, but also on the number of sites from which the content percentage has been taken. As well as if said content is attributed to their respective sources through the correct structuring of their citations, and that in this case it could not be considered plagiarism. However, according to Abad & Sevillano (2018), there is no general rule indicating when or when it cannot be considered plagiarism, or in what percentage it should be classified as a plagiarized document, since the last decision is applied in accordance with the criteria of who reviews the document and the results produced by Turnitin.

Finally, one of the advantages of using a tool to detect possible plagiarism is the review time, according to Abad & Sevillano (2018), it takes a few seconds to analyze 300-page papers. But everything has a price, the cost of using Turnitin depends on the number of users who are allowed access, so for access to 1,000 users there is talk of a

cost of 5,000 euros annually (Abad & Sevillano, 2018). Which means that in Mexican pesos it is equivalent to 135, 024 pesos.

The advantage of this price is that it is a fixed cost, regardless of the number of requests made to the tool.

One of the difficulties that Turnitin encounters is the language, since multiple very different translations can be made from the same text. On the other hand, not having access to all the existing scientific and academic documents results in a major obstacle, in the same way the format of the documents used to compare is not 100% effective, not to mention that it is not all documents are digitized or housed in any database.

3 The practice

In order to strengthen the teaching and learning process of thematic unit number two titled written communication, the development of the program has the competence, which indicates that it distinguishes written communication based on its characteristics, elements and regulations and academic texts (ESCOM, 2020), the topic that was selected from the official study program was 2.5 entitled Academic writing and the critical apparatus subtopic, one of the activities that students must carry out as evidence of learning is academic writing, and As a product, the construction of a research protocol, complying with the format established by ESCOM, when registering its modality of qualification, which has the section of approach to a problem with documentary support, attending to the development of the subject of the critical apparatus.

The experience begins in the exposition of the topic with examples, by the professors, considering how to carry out documentary research, to obtain documentary sources of information, articles of dissemination and articles of scientific dissemination, as well as two systems to make citations within From the body of the document, the systems addressed the APA, as well as the IEEE system, which are given the guidelines to the students, it should be noted that the IEEE system is used by researchers and developers in the engineering area, it is therefore necessary that learn it. The APA system is reviewed, because some research work they will carry out in their academic career will be cited with this system. Another topic addressed was how to perform the critical apparatus, being a topic with a high degree of difficulty for students to do it. This is observed when requesting the construction of the problem statement approach, being a point that must be developed within the format of the research protocol.

The results that have been observed were always similar in a large percentage of similarity of documentary sources of information and without considering copyright, which incurred plagiarism, for the teachers it was an arduous task and it required a long time to do the review. of each job, then an inquiry was made to support the activity, finding Turnitin.

3.1 Participantes

En este trabajo se involucró a 29 estudiantes inscritos en la materia de Comunicación Oral y Escrita (COE) durante el periodo 19-20/2, la cual se imparte en la carrera de Ingeniería en Sistemas Computacionales (ISC) del Instituto Politécnico Nacional (IPN); dicha materia pertenece al primer semestre de la carrera, cabe señalar que en gran porcentaje los estudiantes han sido formados en el área de físico matemáticas, donde ha sido bajo el porcentaje de las materias que han cursado, en donde se fortalece las habilidades de la comunicación oral y escrita.

Cabe señalar que en su trayectoria académica no existe otra materia más que ayude a los estudiantes a escribir su protocolo para que desarrolle su trabajo terminal, y así estar en condición de obtener su titulación curricular. Esto inicia tres semestres antes de que concluya su carrera profesional, los alumnos deberá entregar un protocolo de trabajo terminal para ser aprobado por tres sinodales de diferentes áreas de conocimiento y los cuales darán seguimiento al desarrollo del proyecto tecnológico, siendo el desarrollo de un sistema computacional, luego entonces los estudiantes deben tener desarrolladas las habilidades de la comunicación escrita para la escritura de documentos académicos, siendo el protocolo de trabajo terminal y el reporte de su desarrollo tecnológico.

3.2 Methodological script

After asking the IPN administrator for the authorization of an account in the Turnitin tool and taking a course to learn how to use it, in which he addressed the topic of identifying the tool, its intention, benefits, benefits , as well as the creation of classes and exercises for each class and the registration for the students to make their account, and they themselves upload their activities in each exercise, as well as registering the students by the teachers. In the course that was taken online, it was asked if there were a limited number of classes and the number of students to register for the tool, mentioning the instructor who had no limit on its use.

Once the workshop was over and with the account in the tool, it was decided to create a class with its exercises, to make the invitation to the students.

Once the invitation was made to the students, through the Facebook group titled COE ISC 2020, where the class number and the registration key were shared, as shown in Figure 1, the students began accessing the tool, Only eight of them managed to access and the rest did not, indicating that there was no longer space.

When noticing this action, it was decided to register the students by the teachers, but the result was the same, so we proceeded to report this situation to the administrator, obtaining in response that the IPN has a space of 1,000 students for all levels, being upper secondary, higher and postgraduate. Faced with this situation, the way of working with students had to be modified, integrating them into teams according to their interests in relation to the research topic they decided to work on, however each student should do research on different documentary sources of information, with the intention to strengthen the context of the problem statement.

Fig. 1. Message by Facebook.

The exercises that were opened in class are as follows:

The first exercise, the elaboration of textual records of different documentary sources of information, this with the intention of identifying the priority of consultation of the sources of documentary information.

The second exercise, journal summary worksheets, in order to identify the magazines for student consultation.

The third exercise, mixed worksheets of popular magazines, with the intention of identifying the magazines that the students consult.

The fourth exercise, free worksheets, in scientific magazines.

The fifth exercise, first advance the problem statement in the terminal work protocol format, with textual citations according to the IEEE system.

The sixth exercise, written of the terminal work protocol, with citations and references according to the IEEE.

It is important to mention that from the first exercise, the students already had a defined topic and the research should be from different documentary sources, then the delivery of each individual file should be integrated and delivered as a team.

Each exercise had a period for delivery electronically, in Word or PDF format, which would be published in the Facebook group and from there the teachers downloaded the file to upload it to the tool and obtain the percentage of similarity, the report that is obtained from the tool, is generated immediately, to download and share the result to the members of each team, so that the students identify the similarity of each work, and thus proceed to make the citation correctly from the sources of information thrown by the tool, as observed in figure 2, where the result of a document analyzed by the tool is presented, and figure 3 shows the percentage of similarity as well as the individual result of the percentage of similarity of each research source.

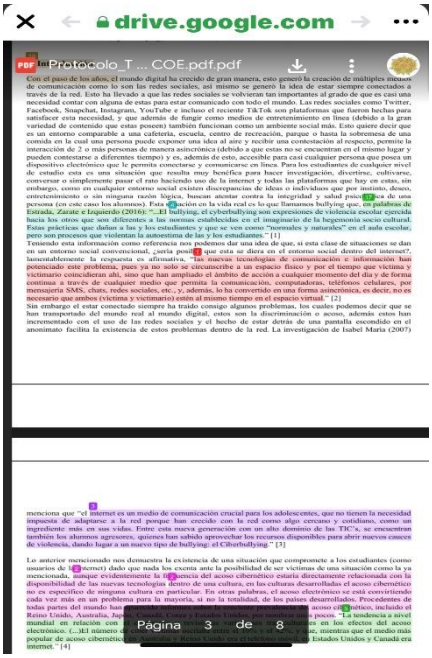

Fig. 2. Document report analyzed by Turnitin.

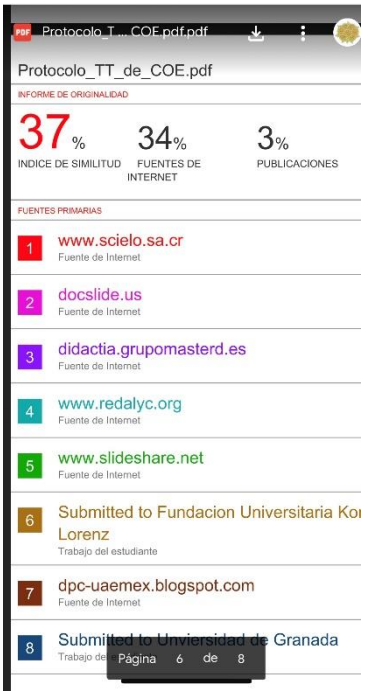

Fig. 3. Percentage of global and individual similarity.

3.3 Context of the practice

The practice was carried out to establish communication with the students through the social network Facebook, generating a group to work on information only on the subject, and to make the process of reviewing the similarity of the works, the Turnitin tool was used is an application that helps you in the detection of plagiarism, through the comparison of academic and professional works with a database, which is included in the tool (Digital Resources, nd).

Turnitin (s.f.) points out that with this tool it identifies unoriginal content through its similarity detection tool, helping in the comparison of the produced content the only database that includes the internet, student works and indexed scientific articles. In addition to assisting in the protection of intellectual property, it provides student tools for peer review and evaluation.

To meet its goals, Turnitin compares jobs submitted across multiple databases using a proprietary algorithm. Among the databases used are their own databases, as well as private academic databases (Turnitin, 2020).

The operation of Turnitin, according to Abad & Sevillano (2018), is as follows:

1. Users upload the documents they want to collate (Abad & Sevillano, 2018).
2. The tool searches for the text strings and compares them with the documents in its databases (Abad & Sevillano, 2018).

3. The user receives as a result the document they have compared and, in case of finding matches, the parts that match are indicated with a different color (Abad & Sevillano, 2018).

4. The final percentage is the percentage of all the text that appears in another document (Abad & Sevillano, 2018).

To know if there is plagiarism or not, according to Abad & Sevillano (2018) depends on the field of study:

• A 20% coincidence in fields like the right one is normal.

• A 3% coincidence in fields such as natural sciences may indicate possible copying.

However, it depends not only on the field of study, but also on the number of sites from which the content percentage has been taken. As well as if said content is attributed to their respective sources through the correct structuring of their citations, and that in this case it could not be considered plagiarism. However, according to Abad & Sevillano (2018), there is no general rule that indicates, when it can or cannot be considered plagiarism, or in what percentage it should be classified as a plagiarized document, since the last decision applies according to the criteria of who reviews the document and the results produced by Turnitin.

4 Results

After having carried out the practice, several results were obtained, being the first that, after having planned the use of the tool with the students, this could not be consolidated, since access to Turnitin is limited for the student population of the IPN For this reason, it was not possible for all the students to register and visualize the similarity result of the documentary sources of information that they had consulted.

It was decided to modify the way of working with the students, since the intention was that it be done individually and it had to be modified to work in a team, however the percentage of similarity in each worksheet could be identified individually delivered by team, and worked individually.

In the first exercise requested, the percentage of similarity was low, because they were textual records of documentary sources of information, here the students were asked to select the sources to obtain information, a proposal made by the teachers.

In the second exercise, the similarity percentage increased on average for the six teams, by 40%, because they were summary sheets.

In the third exercise, mixed sheets were worked and the percentage of similarity obtained by the Turnitin tool increased on average from the six teams to 70%. These works were shared reflecting with the students on the importance of recognizing the work of other authors. .

The fourth exercise they elaborated free worksheets, with magazines of scientific dissemination, the similarity percentage was reduced to 62% on average of the six teams.

The fifth exercise was the first advance of the problem statement, in the terminal work protocol format, with quotations according to the IEEE system, in this advance low percentages of similarity were obtained with 12%, except for one team that obtained the 37%, who were invited to review and correct it.

The sixth exercise consisted of writing the terminal work protocol, with citations and references according to the IEEE, it was reduced to 5% similarity on average of the six teams.

Then it is then declared that the intention of using the Turnitin tool was achieved by decreasing the percentage of similarity of the works considerably, because all the students had the possibility of visualizing the percentage of similarity, this being a challenge for them and attend to ethics as a student.

5 Conclusions

Finally, one of the advantages of using a tool to detect possible plagiarism is the review time, according to Abad & Sevillano (2018), it takes a few seconds to analyze the 300-page work, which is a great saving of time and tasks to be reviewed by the teacher.

On the other hand, the change in attitude on the part of the students when publishing the results, in the Facebook group, and not wanting to obtain high similarity percentages.

It was possible to generate awareness and relevance among the students in the way of giving credit to the authors of the sources of information consulted, managing to avoid plagiarism.

On the other hand, it was obtained, not directly and considerably in the plan, to increase the documentary sources of information and the construction of a critical apparatus argued and cited with sources of scientific dissemination.

The IPN is licensed for use, assigning an administrator, who has the power to add interested teachers, from the institute, to be instructors and each instructor can open classes and invite our students to register and upload their work. However, there is insufficient space to integrate all the students of the three educational levels of the IPN.

References

[1] Abad, J.; & Sevillano, E.: Turnitin y PlagScan: así funcionan los programas antiplagio que ha usado La Moncloa. *elpais.com.* https://elpais.com/politica/2018/09/14/actualidad/1536910999_284317.html. (2018). Accessed: April 2, 2020

[2] Recursos Digitales. *ipn.mx.* https://www.ipn.mx/biblioteca/recursos-digitales/recursos-digitales/. Accessed: April 2, 2020

[3] Turnitin. *Wikipedia.* https://es.wikipedia.org/wiki/Turnitin. (2020). Accessed: April 2, 2020