

## Analyzing how the sobek tool can assist students' writing

DOI: 10.46932/sfjdv4n6-001

Received on: July 28<sup>th</sup>, 2023

Accepted on: August 25<sup>th</sup>, 2023

### **Romário Ferreira de Pontes**

Graduate in Computer Science

Institution: Universidade Federal da Paraíba (UFPB)

Address: Ave. Santa Elisabete, Rio Tinto – PB, CEP: 58297-000

E-mail: romario.pontes@dcx.ufpb.br

### **Thereza Patrícia Pereira Padilha**

Doctor in Computer Science

Institution: Universidade Federal da Paraíba (UFPB)

Address: Ave. Santa Elisabete, Rio Tinto – PB, CEP: 58297-000

E-mail: thereza@dcx.ufpb.br

### **ABSTRACT**

Reading and text production are two important skills for everyone, but it is known that even students in higher education have difficulties how to write high-quality texts. Currently, there are several automatic text analysis tools, also well-known as text mining tools, available on the Internet (paid and free) which can be used to support text production, such as LanguageTool (Zhang et al., 2018), TextAlyser (TextAlyser, 2019), TagCrowd (TagCrowd, 2019), and Sobek (Macedo et al., 2016). These tools usually highlight attention points in a text, present a statistical analysis, suggest connection words, and emphasize locals with repetitive words, turning out a text with higher quality. Thus, this paper presents a case study using the Sobek automatic text analysis tool to investigate its potential to support students in the learning process (reading and production/writing). One group of 26 students enrolled in 9th grade from a middle school in the Paraíba state, Brazil, participated in writing activities using the Sobek tool. From the experiments, the results have shown an upgrade in text writing skills, especially showing a vocabulary more diversified and text structures, which are discussed in this paper.

**Keywords:** educational software, automatic text analysis, writing skill, education.

### **1 INTRODUCTION**

Nowadays, the technological development has helped to carry out activities in several fields of society. In the educational field, for example, educators have had a wide range of resources to teach students, providing more meaningful learning and increasing the dynamism in classrooms. The technological advances in educational tools have supported knowledge construction and acquisition, allowing occur it at any time and space. The educational tools can support writing, especially, in what cognitive aspects of writing are concerned in the learning process. So, educations tools that are associated with planning, generating ideas, organizing ideas, and goal-settings are welcome.

Effective writing is a skill that is grounded in the cognitive domain. It involves learning, comprehension, application and synthesis of new knowledge (Defazio et al., 2010). Writing even though reading skills are considered fundamental factors for the insertion of human being into the labor market. The reading provides the access to information, vocabulary expansion, development of critical thinking, and interest in the search for knowledge on various subjects that can boost her/his social relationships. Writing, on the other hand, is directly related to the gathering of words, organization and development of ideas, formation of critical thinking, and creation of argumentative/narrative essay writing. Thus, these questions directly reflect the formation of active, reflective, and critical human being.

With the emergence of computers and the Internet, the production of documents, such as reports, email messages, and papers, has become common the use of text edition programs. Consequently, a huge amount of structured and unstructured data has been generated and stored in supercomputers. As text data processing is a complex and challenging problem, the text mining area was created to focus on the analysis/extraction of information through methods capable of dealing with a large number of words and structures in a natural language and to support the activities of comprehension and text production.

Text mining is viewed as a multidisciplinary field including knowledge of areas such as Informatics, Statistics, Linguistics, Cognitive Science, and others. It is one process of extracting previously unknown, understandable, potential and practical patterns or knowledge from the collection of text data (Zhang et al., 2015). One of the text mining techniques is called of **information extraction** that analyses and identifies automatically main terms and phrases in texts, and their relationships. Extraction infers relationships between terms/facts, places, dates to help people to improve their text production (Mooney & Nahm, 2003).

Currently, there are several text mining tools, also known as automatic text analysis tools, available on the Internet (free and paid), which can be used to support text production, such as Languagetool (Zhang et al., 2018), TextAlyser (TextAlyser, 2019), TagCrowd (TagCrowd, 2019), and Sobek (Macedo et al., 2009). These tools usually signal unclear text parts, present a statistical analysis, identify the connection between words or terms, and highlight the repetitive words. Thus, they help students to produce text clearer and more concise.

LanguageTool (<https://languagetool.org/>) is an open-source proofreading software (spelling and grammar checking tool) for English, French, German, Polish, Russian, and other languages. It finds many errors that a simple spell checker cannot detect. For that, an user can just provide his/her text in the space provided on the platform, choose the language (English is the default), and automatically sentences with possible errors are highlighted. It is possible to use a free version (until 20k characters) and paid. TextAlyser (<https://seoscout.com/tools/text-analyzer>) is another free online tool that aims to identify the keywords of a given text. Initially, this tool counts the used terms in the text and presents their statistics

(frequency). It also shows an index related to “readability”, obtained from the size of the sentences and statistics. Sobek (<http://sobek.ufrgs.br/index-en.html>) is also another free tool that extracts the most important concepts of the text in different formats (.pdf, .doc, or .txt) and presents their interconnections of visual mode (graphs) to help the writing and the summarization. It is possible to use this tool on the Desktop (Linux, Windows, or Mac OS) and online versions. The Sobek’s goal is provide a concise view of a text (graphical representation), highlighting clues about problems, or about the quality of a text, that can be recognized promptly.

Table 1 shows a brief comparison of these three tools emphasizing some features.

Table 1: Comparative among Sobek, TextAlyser, and LanguageTool tools.

Features	Sobek	TextAlyser	LanguageTool
Online	✓	✓	✓
Terms counting	-	✓	✓
Relevant terms highlight	✓	-	-
Terms frequency	✓	✓	-
Relationship among terms	✓	-	-
Graphic visualization of terms and their relationships	✓	-	-

Source: Data built from our analysis.

Analyzing the existing features of these tools, we chose Sobek for this work because it presents results in a visual mode allowing an easier interpretation and works for all the operating system’s computer users. In addition, several studies have been showing Sobek as a good text production tool (Reategui, et al., 2011), (Lovato, 2015), and (Azevedo et al., 2011).

This paper presents the results of the use of the Sobek automatic text analysis tool in order to investigate its potential to assist students in the learning process that involves reading and, mainly, writing skills. So, this tool was presented to a sample of 26 9<sup>th</sup> grade students at the Ulisses Mauricio de Pontes Middle School, in Sertaozinho city, Paraiba, Brazil. This school was chosen because, according to Basic Education Development Index (Ideb) collected in 2019, only 26% of the students (9th graders) learned what was supposed to in Portuguese language class (text reading and writing) which is considered below-average performance. The Ideb is a metric that the Brazilian government calculated based on student learning in Portuguese and mathematics (Brazilian exam called “Prova Brasil”) and approval rate (Ideb, 2019).

This present paper is organized as follows: Material and Methods section presents the used methodology for the experiments. In the Results and Discussion section, observations from the meetings with the students and Sobek tool are described. Finally, the final considerations are presented.

## 2 MATERIALS AND METHODS

The proposal was to analyse the text mining tool Sobek as a intelligent resource to get ideas about the texts written by students. To understand better how the Sobek tool could help in text production, 26 students who were enrolled in 9<sup>th</sup> grade at Ulisses Mauricio de Pontes Elementary School, Sertaozinho, Paraiba – Brazil, accepted to participate in this research. First, a survey was applied to them to evaluate how much they know about computing (e.g. knowledge about operating system, basic computer, Internet, Microsoft Word application) and text production (e.g. grammar, clarity, content, style, critical thinking, and others). In addition, 4 in-person meetings were planned, 60 minutes each, to observe and practice some writing activities with and without the Sobek tool. The schedule of the meetings can be seen below:

- **Meeting 1:** presentation of the basic computer concepts and features of the editing software (Microsoft Word);
- **Meeting 2:** reading, text comprehension, and identification of terms and relationships, creating a graph with terms and relationships manually (conceptual map);
- **Meeting 3:** knowing the functionalities of the Sobek text analysis tool;
- **Meeting 4:** reading, comprehension, and text production using the Sobek tool.

After all the meetings, one Portuguese teacher [School\_Name] evaluated all written materials by students (with and without the use of the Sobek tool) and compared them to verify the gain of quality when students use automatic text analysis tool in their activities.

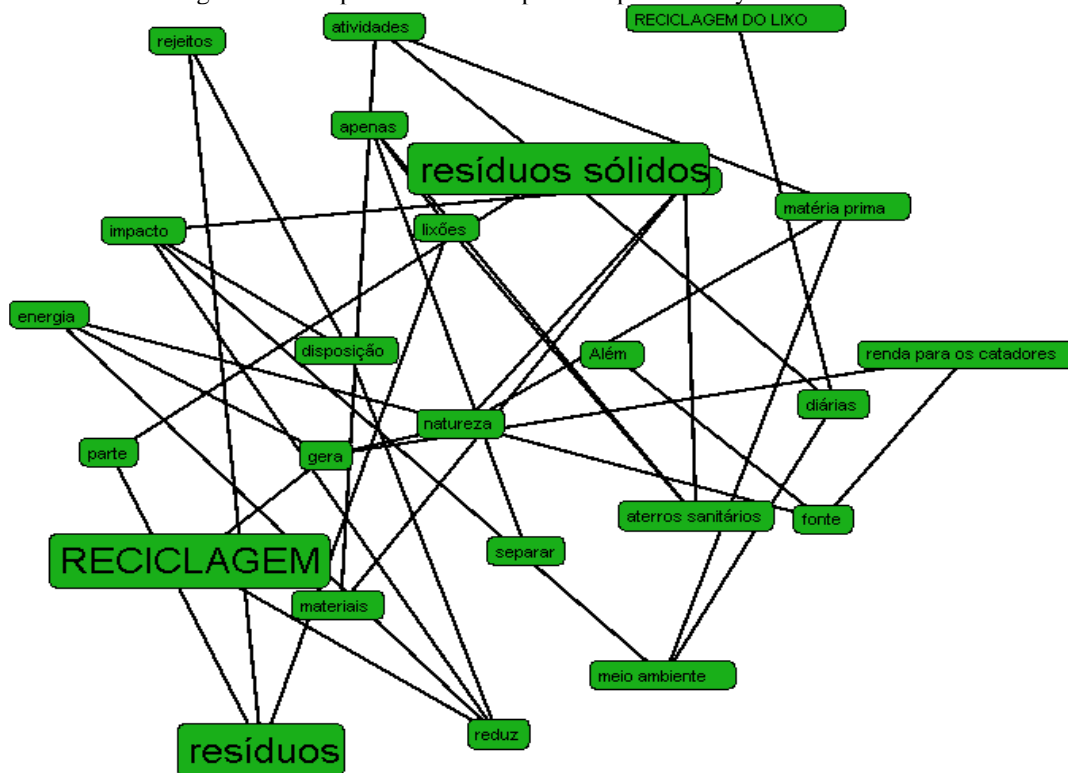
## 3 RESULTS AND DISCUSSION

Based on the methodology cited, the survey results indicated that a significant proportion of students lacked knowledge about approximately 60% of the topics covered. Additionally, the survey revealed that some students had not yet utilized valuable features of Microsoft Word software, such as finding synonyms for words when writing and formatting. The main reasons were that most of them did not have computer/laptop/tablet at home and at school they had just a few opportunities to use computers. The observations and/or results achieved in each Meeting are presented below:

- **Meeting 1:** students learned how to use basic functions of a computer (to start/shut down, access the Internet, open/close applications). In word processing software (Microsoft Word), they learned how to type a text, especially, how to add accentuation, punctuation, formatting and the use of synonyms;
- **Meeting 2:** students wrote an essay about “*What’s the importance of waste recycling?*” using the Word software. After that, students read their essays trying to identify important words, track possible relationships among them, and then created graphical representations (conceptual maps);

- **Meeting 3:** students had an opportunity to know the main functions of the Sobek tool, access its website, and understand its proposal. In general, all students were engaged with this pedagogical resource that can help them in Portuguese classes, improving their writing skills;
- **Meeting 4:** students used the Sobek tool to write an essay about the “*Environment*”, which resembles the one mentioned in M2. On this occasion, they explored the available functions to improve their texts. After that, a Portuguese teacher compared both essays (output from Meeting 2 and 4) written by each student. In her view, all essays developed in the Meeting 4 were better than the previous ones observing grammar, vocabulary, clarity, content, style, and critical thinking. In general, students were attentive to the use of synonyms (expressing more vocabulary), the connections among sentences (linking words), and consequently, the content. Figure 1 shows an example of a conceptual map extracted from a final text produced by one student in the Sobek tool.

Figure 1: Example of one Conceptual Map created by the Sobek tool.



Source: Data extracted from our experiments.

By looking at this conceptual map, it was possible to say that it had important concepts related to the central topic of the assignment. The biggest rectangles represent the most important terms, such as solid waste and recycling (in Portuguese: “resíduos sólidos” and “reciclagem”, respectively). And the links represent the existing connections among them. In general, the Sobek tool provided positive and/or negative clues about a text, enabling, for example, the identification of gaps about the subject and then

the need for further exploration of a given topic in order to create connections of paragraphs. As there is no isolated concept, it shows that the information are all related.

#### **4 CONCLUSION**

This paper presented results of a study involving the use of the Sobek tool to help students in text production. In total, we had 26 participants (9<sup>th</sup> graders) who had weak computer and writing skills. For this reason, before the writing experiments, we initially taught some basic computer lessons such as how to start and shut down a computer, access word processing software, and visit the Sobek website.

By presenting the visualized result using the text mining Sobek tool, it was possible to analyse the relationship between the words (vocabulary) in order to understand the overall content structure. It brings a qualitative evaluation of written material produced by. So, tools like Sobek can be viable to the writing process, identifying weak and strong points. About these experiments, the Portuguese language teacher thought that the written materials by participants using the Sobek tool were more comprehensive if compared to when they used only the word processing software. According to her evaluation, the highlighted words (most relevant terms - keywords) in the text and the graphic showing the main concepts and their relationships have helped students to understand the main point and, consequently, improve how they were writing. The advances were the reduction of the repetitive words, the need of looking for synonyms (more vocabulary) and the use of the linking words, according to the teacher observation. Therefore, it is possible to conclude that the use of text mining tools, such as Sobek, can help students in text production and engagement in the classroom's activities showing how students could benefit from automatically seeing summaries of their writings.

## REFERENCES

- Azevedo, B. F. T.; Behar, P. A.; Reategui, E. (2011). Qualitative Analysis of Discussion Forums. *International Journal of Computer Information Systems and Industrial Management Applications*. ISSN 2150-7988 Volume 3, 671-678.
- Defazio, J.; Jones, J.; Tennant, F.; Hook, S. A. (2010). Academic literacy: The importance and impact of writing across the curriculum – a case study. *Journal of the Scholarship of Teaching and Learning*, Vol. 10, No. 2, 34 - 47.
- Ideb (2019, August 1). <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/ideb>.
- Lovato, G. (2015). *Aplicação da Mineração de Textos na Análise ne Produções Textuais*. Trabalho de Conclusão de Curso em Ciência da Computação, Universidade de Caxias do Sul.
- Macedo, A.; Reategui, E.; Lorenzatti, A.; Behar, P. (2009). Using Text-Mining to Support the Evaluation of Texts Produced Collaboratively. *Conference: Education and Technology for a Better World, 9th IFIP TC 3 World Conference on Computers in Education, WCCE 2009, Bento Gonçalves, Brazil, July 27-31*.
- Mooney, R. J.; Nahm, U. Y. (2003). "Text Mining with Information Extraction". In *Proceedings of International Midp Colloquium*, 141-160.
- Reategui, E.; Klemann, M.; Epstein, D.; Lorenzatti, A. (2011). Sobek: a text mining tool for educational applications. *International Conference on Data Mining, Las Vegas, Nevada, USA*. 59-64.
- TagCrowd, (2019, September 7). Retrieved from URL <https://tagcrowd.com/>.
- Textalyser Textalyser, (2019, September 7). Retrieved from URL <https://textalyser.net>.
- Zhang, Y.; Chen, M.; Liu, Liu. (2015). A review on text mining. *6th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, pp. 681-685, doi: 10.1109/ICSESS.2015.7339149.
- Zhang, W.; Lin, Z.; Xiao, G.; Chen, J.; Wang, J.; Jiang, Y. (2018). *LanguageTool Proofreading Rules Evolution and Update*, 95-100. <https://doi.org/10.1145/3226116.3226134>.