

ANALYSIS AND COMPARISON OF ONLINE PLATFORMS FOR PLAGIARISM: COMPARING THE EVALUATION SYSTEM OF THE MINISTRY OF EDUCATION AND SCIENCE OF NORTH MACEDONIA WITH OTHER ONLINE SYSTEMS

Enes BAJRAMI^{1*}, Ermira MEMETI¹, Agon MEMETI¹

¹*Department of Computer Science, Faculty of Mathematical and Natural Science, University of Tetova*

^{*}*Corresponding author e-mail: e.bajrami3202009@unite.edu.mk*

Abstract

Plagiarism is when another author's language, thoughts, ideas, or expressions are presented as one's own. Plagiarism is defined variously in educational settings depending on the institution. Plagiarism is regarded as a form of academic cheating. The main purpose of this paper is to analyze systems and discover the accuracy of systems for plagiarism, using documents of different sizes. It is critical to raise researcher knowledge of plagiarism of words, ideas, and graphics in order to avoid unethical writing methods. Updating author instructions and informing plagiarists of academic and other consequences of unethical behavior should be done by consulting linked papers. Most instances of redundant and "copy-and-paste" writing are said to be caused by a lack of innovative thinking and weak academic English skills. Plagiarism detection software mostly depends on text similarity reporting. Manual checks, on the other hand, are required to detect improper referencing, copyright breaches, and poor English writing.

Keywords: Plagiarism, Academic, Text, System, Paper, Detection

1. Introduction

The problem of plagiarism has recently increased because of the digital era of resources available on the World Wide Web. Plagiarism detection in natural languages by statistical or computerized methods has started since the 90s, which is pioneered by the studies of copy detection mechanisms in digital documents [1][2]. Earlier than plagiarism detection in natural languages, code clones and software misuse detection has started since the 1970 by the studies to detect programming code plagiarism in Pascal and C [3], [4]-[7]. Algorithms of plagiarism detection in natural languages and programming languages have noticeable differences. The first one tackles different textual features and diverse methods of detection, while the latter mainly focuses on keeping track of metrics, such as number of lines, variables, statements, subprograms, calls to subprograms, and other parameters. During the last decade, research on automated plagiarism detection in natural languages has actively evolved, which takes the advantage of recent developments in related fields like information retrieval (IR), cross language information retrieval (CLIR), natural language processing, computational linguistics, artificial intelligence, and soft computing. In this paper, a survey of recent advances in the area of automated plagiarism detection in text documents is presented, which started roughly in 2005, unless it is noteworthy to state a research prior than that [8], [9]-[12]. The main focus of this paper is the analysis and comparison of platforms for plagiarism. As a basic platform we have taken the system of the Ministry of Education and Science of the Republic of North Macedonia where we have compared the same with other platforms for the accuracy and speed of detecting the same words from articles on the web.

2. System for plagiarism detection and analysis of the Ministry of Education and Science of the Republic of North Macedonia

This system provides an easy and efficient way of detecting documents that contain plagiarized parts from already published and presented documents. The system's goal is to provide an easy and intuitive interface for uploading documents. The students and researchers/scientists can easily upload their homework, bachelor's thesis, master's thesis, doctoral thesis (dissertation) and other published papers and documents. The system also provides a mechanism for comparing the uploaded documents with all the other documents that are already present in the system and to measure their originality, i.e., to detect if any of the content is already published [13]. The ministry system is built with Java programming language, JavaScript Framework is used Prototype version 1.7, is using Debian Operating System and Apache for Web Server, also using JavaScript Libraries Script Aculo and jQuery, for UI Bootstrap.

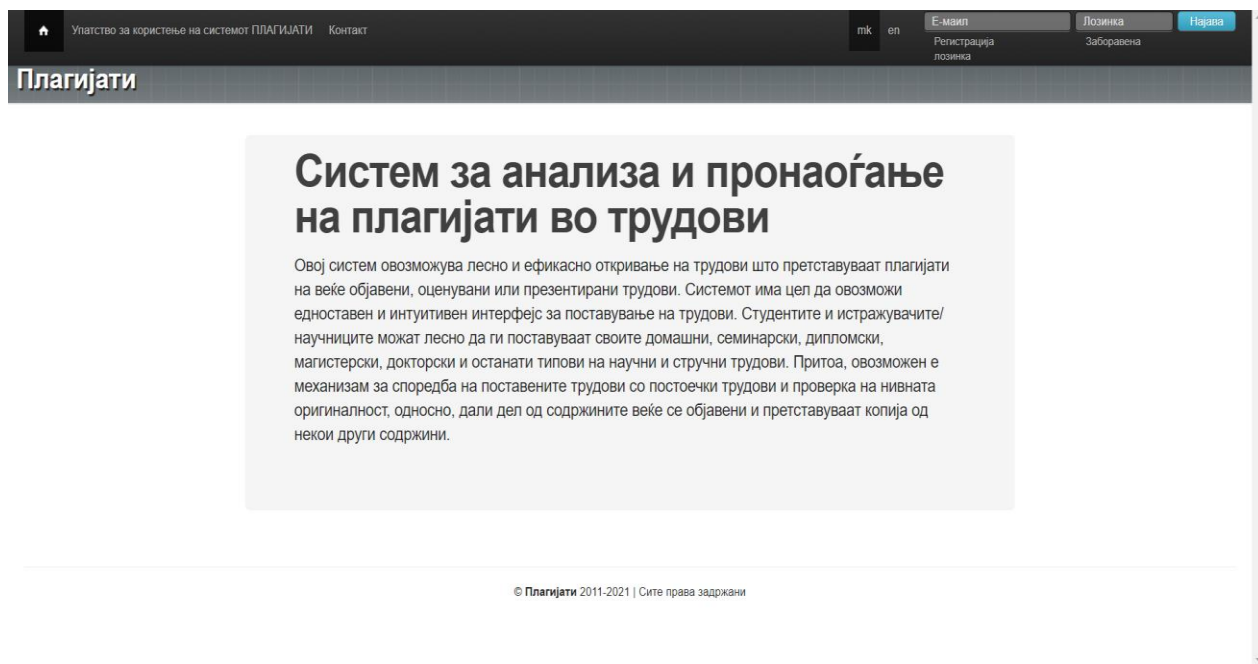


Fig 1. Plagiarism System Homepage of Ministry

3. Online plagiarism tools

Free plagiarism detection tools along with the website address are mentioned in below tables. Software's which are freely available online are mentioned in Table 1 and user can use it by simply open the link.

Benefits of using anti-plagiarism detection tools:

- With the help of plagiarism detection tools researcher can find out if any other person copied his/her research writing.
- It helps to improve writing skills as some plagiarism tools check grammar also.
- It filters the duplicate content.
- These tools are useful to avoid intellectual property issues.
- It promotes effective and efficient research.
- These software's are useful for research scholars, website creator, and publisher.

- Plagiarism tools allow to access multiple databases.
- It allows to detect any incidents of plagiarism and helps to take appropriate action.
- It helps the researcher to properly cite references
- It creates awareness of plagiarism among research scholars, faculty members and help them to have successful academic careers in future by avoiding plagiarism.

Limitations of detection tools:

“The drawbacks of detection tools are:

- Inability to distinguish correctly cited text from plagiarized text.
- Books are typically not searched by these services.
- Detect plagiarized words, not plagiarized thoughts or ideas.
- Inability to process textual images for similarity checks” [14][15]

Table 1. Plagiarism detection tools freely available online

Online Software Name	URL	Method of checking	Advantages and Disadvantages
Plagiarism Checker[16]	http://smallseotools.com/plagiarism-checker/	The user must paste the content into the provided box and then click the huge green "Check for Plagiarism!" button. It will examine the material and compare it to online sources before generating a report.	Adv: This application can help you find the original source of plagiarized text you've copied from the internet. It has the ability to verify content line by line. Returns answers very quickly. Dis: It can only allow 1000 words in a single search at a time.
Duplichecker[17]	https://www.duplichecker.com/	The user must paste or upload the content file and then click the "search" button to check for plagiarism. It will analyze the content and compare it to web sources before generating a report.	Adv: It's highly valuable because it displays all of the source websites from which the text was copied, and researchers can utilize it to provide accurate references. Returns answers very quickly. Dis: You can check a maximum of 1500 words in a search at the same time.
Ministry of Education and Science of North Macedonia[13]	http://plagijati.mon.gov.mk/	First the user has to register, then he has to attach the file to the system to get the result.	Adv: the system is very accurate, allows you to place files of different sizes and there is no limit for words in search. supported file formats [lit, xlsx, json, docx, csv, pdf, epub, odt, txt, doc, html, xls, xml] Dis: Response time is long for result; The user needs to be registered to check the result

4. Related work

Modern studies on plagiarism in books and literature have been reviewed and your discovery reveals that wrongdoing around the world has been growing to create uncontrollable hands. Gorman (2008) has described that people use the literature that they find on the internet without the fear of copyright infringement using a direct copy source. M.P. Satija (2015) wrote a article with the title “Preventing the plague of plagiarism”. In this article it was mentioned that more than 17000 plagiarisms were mentioned in the British universities including Oxford.

The funny thing was that even in the Harvard University plagiarism was present (Tribune,2013,p.14).

Most plagiarism cases in North Macedonia come to the fore due to ignorance of the concept of copying, copyright and intellectual property rights. The same problem is also present in India.

Hoorn, et al explained that "there is no" awareness of various issues "related to copyright and intellectual property rights among participants, law enforcement agencies, and practitioners around the world as research communities, the study revealed that 30 percent of respondents were unaware. "First-time copyright holders of their research papers, and 26 percent of respondents showed a low interest in copyright issues in their research articles" clearly showing that researchers are doing research or making art in a specific area but with a low interest in preventing copying even if you own the same copyrighted work. by someone else. T.A. Abhinandan analyzed “69 retrospective papers by Indian writers, finding that the national retracement rate of 44 per 1,00,000 pages was higher than the world average of 17 to 1,00,000 in total withdrawals”

According to [18], Dupli Checker is one of the best free plagiarism detector tools on the internet. It does not have an extravagant interface, but it takes care of business. It allows you to reorder the content and also check for plagiarism.

5. Interpretation of research results

As seen in the section below we have presented some diagrams with results where we have compared systems with different word sizes, each of the system has shown different results both in time measurement and in percentage of plagiarism. The diagrams are presented in the form of charts with all the needed information.

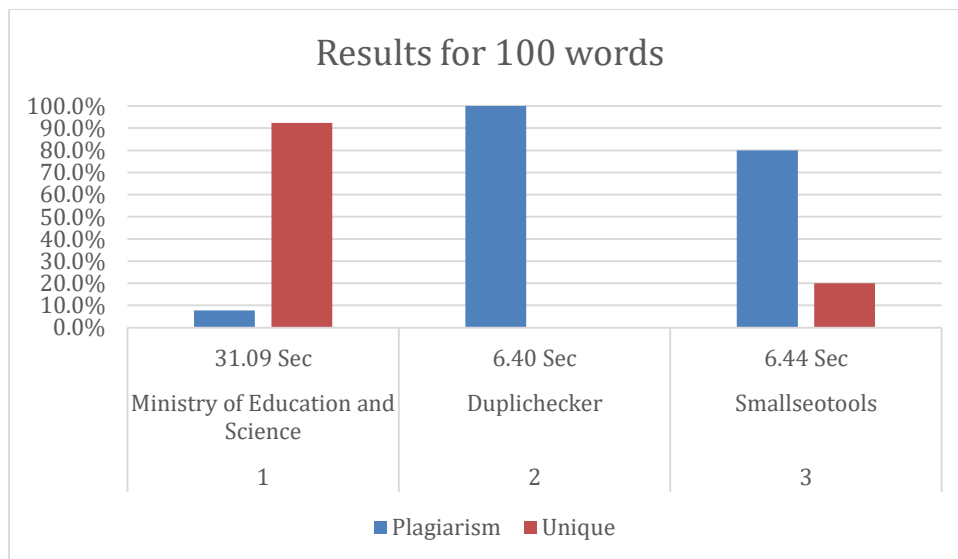


Fig 2. Results for 100 words

As shown in the diagram above, we have user three different systems and the same length of words. The length of words in this diagram is 100.

The first system is “Ministry of Education and Science” and from the results we concluded that for 100 words it took the system 31.9 seconds to detect the plagiarism. From results you can see that only 10% of plagiarism was detected.

The second system is “Duplichecker” and from the results we concluded that for 100 words it took the system 6.40 seconds to detect the plagiarism. From results you can see that 100% of plagiarism was detected.

The last system is “Smallseotools” and this system took 6.44 seconds to proceed 100 words to find the plagiarism. From results you can see that only 80% of plagiarism was detected.

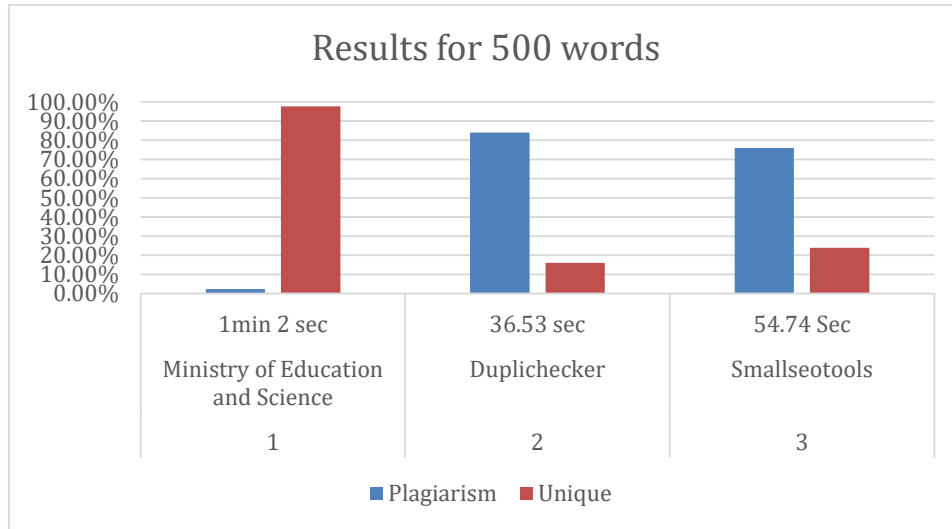


Fig 3. Result for 500 words

The same three systems were used. But this time we have used a length of 500 words. Results are:

- The first system – for 62 seconds found only 2% of plagiarism
- The second system – for 36.53 seconds found 85% of plagiarism
- The third system – for 54.74 seconds found 75% of plagiarism

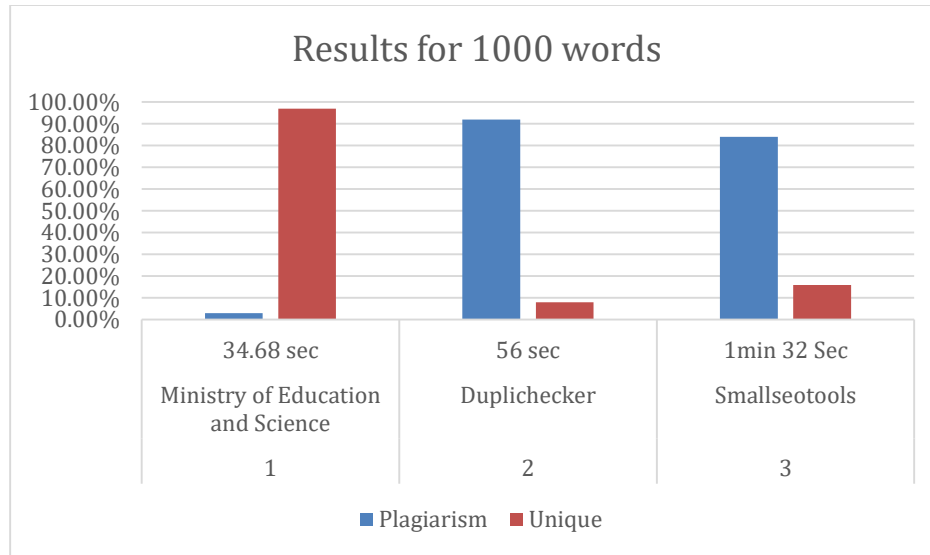


Fig 4. Result for 1000 words

The final experiment was with 1000 words and results are:

- The first system – for 34.68 seconds found only 3% of plagiarism
- The second system – for 56 seconds found 91% of plagiarism
- The third system – for 92 seconds found 83% of plagiarism
- From the three of systems that we have used, according to all the experiments that we have made, the best system is the second one (Duplichecker) and the worst one is the first one (Ministry of Education and Science)

6. Conclusions

In this paper, both authors have analyzed the platforms for plagiarism, in particular the system of the Ministry of Education and Science, how it is built and how it works. They have compared the same system with 2 other online platforms, where we can mention that both platforms have shown better results than the system of the Ministry of Science and Education. Besides our thorough research on this paper, there are still some gaps left that would be recommended for future research, like: result for more than 1000 words, system accuracy, the origin of plagiarism.

References

- [1]. S. Brin, J. Davis, and H. Garcia-Molina, "Copy detection mechanisms for digital documents," in Proc. ACM SIGMOD Int. Conf. Manage. Data, New York, 1995, pp. 398–409
- [2]. N. Shivakumar and H. Garcia-Molina, "SCAM: A copy detection mechanism for digital documents," in D-Lib Mag., 1995
- [3]. A. Parker and J. O. Hamblen, "Computer algorithms for plagiarism detection," IEEE Trans. Educ., vol. 32, no. 2, pp. 94–99, May 1989.
- [4]. K. J. Ottenstein, "An algorithmic approach to the detection and prevention of plagiarism," SIGCSE Bull., vol. 8, no. 4, pp. 30–41, 1977.
- [5]. L. D. John, L. Ann-Marie, and H. S. Paula, "A plagiarism detection system," SIGCSE Bull., vol. 13, no. 1, pp. 21–25, 1981
- [6]. G. Sam, "A tool that detects plagiarism in Pascal programs," SIGCSE Bull., vol. 13, no. 1, pp. 15–20, 1981.
- [7]. K. S. Marguerite, B. E. William, J. F. James, H. Cindy, and J. W. Leslie, "Program plagiarism revisited: Current issues and approaches," SIGCSE Bull., vol. 20, pp. 224–224, 1988

- [8]. Z. Ceska, "The future of copy detection techniques," in Proc. YRCAS, Pilsen, Czech Republic, pp. 5–10.
- [9]. P. Clough, "Plagiarism in natural and programming languages: An overview of current tools and technologies," Dept. Comput. Sci., Univ. Sheffield, U.K., Tech. Rep. CS-00-05, 2000
- [10]. P. Clough, (2003) Old and new challenges in automatic plagiarism detection. National UK Plagiarism Advisory Service. [Online]. Available: http://ir.shef.ac.uk/cloughie/papers/pas_plagiarism.pdf
- [11]. H. Maurer, F. Kappe, and B. Zaka, "Plagiarism—A survey," J. Univ. Comput. Sci., vol. 12, pp. 1050–1084, 2006
- [12]. L. Romans, G. Vita, and G. Janis, "Computer-based plagiarism detection methods and tools: An overview," presented at the Int. Conf. Comput. Syst. Technol., Rousse, Bulgaria, 2007
- [13]. <http://plagijati.mon.gov.mk/en/>
- [14]. Avoiding Plagiarism in Research through Free Online Plagiarism Tools, 2015 4th International Symposium on Emerging Trends and Technologies in Libraries and Information Services
- [15]. Jaya, P. A., "Seminar report on plagiarism detection techniques". Available on: <http://seminarprojects.com/Thread-seminar-report-onplagiarism-detection-techniques>
- [16]. <https://smallseotools.com/plagiarism-checker/>
- [17]. <https://www.duplichecker.com/>
- [18]. Nandkishor Balu Gosavi, "Anti-Plagiarism Tools: A Review", World Journal of Humanities and Social Sciences Vol 1, Issue 1, March 2021