Survey Paper on Winnowing Algorithm

Amna Islam[1], Arpit Mishra[2], Er. Anuj Singh[3]

[1]UG student of Department of Computer Science and engineering, Shri Ramswaroop Memorial College of Engineering and Management Lucknow, Uttar Pradesh, India

[2] UG student of Department of Computer Science and engineering, Shri Ramswaroop Memorial College of Engineering and Management Lucknow, Uttar Pradesh, India

[3] Associate Professor, Department of Computer Science and engineering, Shri Ramswaroop Memorial College of Engineering and Management Lucknow, Uttar Pradesh, India

---***---

Abstract - The following survey paper focuses on the adoption of the Winnowing algorithm for developing a plagiarism tool for learning environments. It is worth mentioning that the Winnowing algorithm is especially effective in identifying probable matches if the same sequences of citations, hypotheses, or legends are given in the student's assignment. There are a number of researches that have been conducted to try to determine how the Winnowing algorithm can be used in detecting plagiarism in varying contexts like text documents, source code and even final project titles among others. The existing algorithm has been improved further by modifying it and looking into the inclusion of k-gram values and comparing various stemmers for plagiarism purposes. These results reinforce the importance of Winnowing in highlighting areas of similarity and even plagiarism in academic works. The purpose of this paper is to give the most complete description of the algorithm Winnowing as it is used in the context of plagiarism detection within educational institutions and to the development of the concept of academic honesty and originality among learners.

1.INTRODUCTION

Poor referencing is where a person uses another individual's work without citing the source as defined in almost all educational spheres. In the recent decades, the use of internet has been at the fore front therefore, students end up copying from one another hence cheating in the classroom. At the same time, the need to develop reliable and effective plagiarism detection systems has become inevitable. One of the techniques aggressively being explored is the use of the Winnowing algorithm which is known as the most efficient way to compare similarities between documents.

This survey paper addresses the performance of using the Winnowing algorithm for the implementation of the plagiarism checker in the educational environment. The Winnowing algorithm created for the detection of similarities in source code can be used to detect documents' plagiarism of any other piece of text. The algorithm may also be used to determine the percentage

similarity between different texts by comparing fingerprints and hash values derived from the documents and it can potentially help in detecting instances of plagiarism.

Many researchers have used the Winnowing algorithm to reveal the similarity between academic documents (text documents), technical files and the final project names. Some prominent works done on Winnowing algorithm in recent years include paid attention to increasing the intensity of this algorithm by applying modifications like k-gram value or how different stemmers deliver the algorithm performance. These improvements focus on developing the algorithm to more precisely and quickly identify similarities in documents. Furthermore, there are investigations on ability of the Winnowing algorithm to detect plagiarisms in contrast with other forms of plagiarism detection.

The objective of this survey paper is to review the recent work that is concerned with the use of the Winnowing algorithm to detect plagiarism in educational settings. In order to achieve this, this paper will seek to make comparisons using combined data from other studies in an effort to determine the usefulness of the Winnowing algorithm for detecting plagiarism. Moreover, the survey would analyze how the algorithm plays a role in encouraging students to meet the standards of academic integrity and originality in classroom environments.

This survey paper aims to merge into the discussion of the plagiarism detection problem in educational institutions and contribute to finding the solution by studying the use of the Winnowing algorithm. This paper will try to assist readers in understanding the implementation of a plagiarism checker using the Winnowing algorithm in classrooms by discussing the existing research and extracting important information from those works

2. LITERATURE SURVEY

Plagiarism is a significant concern in educational settings, necessitating the development of effective detection systems. The Winnowing algorithm has emerged

as a promising tool for identifying similarities in documents and detecting plagiarism. Several studies have explored the application of the Winnowing algorithm in detecting plagiarism across various contexts, including text documents, source code, and final project titles.

[1] Aljohani & Mohd's (2014) paper focuses on cross-language plagiarism using Arabic and English as well as the Winnowing algorithms. More specifically, the algorithm is interpreted in the context of this research as a tool for detecting plagiarism in texts written in Arabic and English. In other words, the study will seek to improve the algorithm's performance of generating fingerprints and comparing textually similar content across languages towards improving the odds of identifying plagiarism. This paper benefits the discipline of computer science in that it tackles the problems with identifying plagiarism that emanate from multilingual settings, particularly the translation between Arabic and English. The Winnowing algorithm which is known for its efficiency in identifying the similarities between documents and plagiarism detection is used in a cross-language mode of operation to address the marked differences of the two languages; Arabic and English used in the research. Applying the above algorithm in this study will improve and advance the plagiarism detection in different languages. The research underlines the importance of creating specific strategies that might be applied to solve the complexity of cross-language plagiarism detection for contributing to the progress of information technology or computational linguistics, including the Winnowing algorithm. The findings of the study will be expected to offer information on the efficiency of the winnowing algorithm in identifying plagiarism in Arabic and English documents. The study is designed to explain the cross-language plagiarism detection in more detail and seek to utilize the capabilities of the algorithm to provide concrete recommendations for detecting similarities and possible cases of plagiarism in multi-lingual texts.

[2] This paper by Velásquez & Taylor (2014) aim at identifying tools for external plagiarism detection within the DOCODE system. Million states that the algorithms provided by DOCODE help identify different plagiarism cases through various approaches. In particular, the Web-based external plagiarism detection with the help of such sophisticated algorithms and the description of the system's effective application of the mentioned algorithms are the major benefits of the study. They also laid emphasize on the application of advanced methods and algorithms to prevent external plagiarism. Coordination of advanced technologies in the DOCODE system will enhance successes of researchers and experts to identify plagiarism from external sources efficiently. It is noteworthy to use highly complicated formulas and all kinds of technologically-inclined means to maintain the moral standards and credit-worthy compositions in the academic works. It is believed that the conclusion of this study will be Useful in establishing and determining the effectiveness of an external plagiarism detection tool within the DOCODE system. As the result of the study, the audience will be shown how the system uses the algorithm that allows checking plagiarism coming from external sources, and thus the work of the research adds to the existing arsenal of efficient and the most effective means to enhance academic integrity and originality.

Irjet template sample paragraph Irjet template sample paragraph. Irjet template sample paragraph Irjet template sample paragraph Irjet template sample paragraph Irjet template sample paragraph. Irjet template sample paragraph Irjet template sample paragraph Irjet template sample paragraph Irjet template sample paragraph Irjet template sample paragraph. Irjet template sample paragraph Irjet template sample paragraph Irjet template sample paragraph Irjet template sample paragraph Irjet template sample paragraph.

Irjet template sample paragraph, Irjet template sample paragraph .Irjet template sample paragraph. Irjet template sample paragraph

[3] The study by Nugroho (2017) explores the effects that the stemming algorithm of Nazief and Adriani in the identification of plagiarism for the Winnowing algorithm with special reference to the Indonesian language. The aim of the study is to evaluate the performance of Winnowing algorithm when it is supplemented with the implementation of STEM algorithm of Nazief –Adriani in determination of plagiarism in Bahasa Indonesia. Specifically, the signals of the Indonesian language will allow to use the algorithms effectively to improve the efficiency and accuracy of plagiarism detection for Indonesian students. This work will advance knowledge in the computer science discipline through offering understanding into enhancing stemming algorithms which is in this case the manner in which the algorithm such as the Nazief-Adriani might be integrated with other plagiarism detecting algorithms like Winnowing to improve its performance in detecting issues of plagiarism in Bahasa Indonesia forms.

[4] The research by Rabbani and Karnalim (2017) seeks to propose a machine learning-based approach to facilitate the detection of source code plagiarism. The study incorporates three classification algorithms; Naïve Bayes, K-Nearest Neighbor (KNN), and AdaBoost Meta-learning algorithm, to improve detection of plagiarized values in source code. The utilization of state-of-art machine learning techniques is therefore an important goal for the study in the effort to minimize the cases of plagiarism that are evident in programming assignments and source code repositories. This research also plays its part in enhancing the continuous improvement of the methodologies of plagiarism detection through use of complex machine learning techniques specifically

designed for source codes. The integration of the machine learning algorithms in source code plagiarism identification is fundamental as it contributes towards the development of better performance of the automated systems. The goal of the study is to develop an efficient, multi-modal classification system where different algorithms are incorporated in order to compare text strings and potentially flag plagiarized material written in a variety of programming languages. The study also appreciates the need to encourage the usage of machine learning techniques to tackle the contemporary issues of source code plagiarism detection which has emerged to be a complex task. The findings of this study are hoped to give insights about the effectiveness of using the information gain machine learning approaches to detect the source code plagiarism. Thus, the research is beneficial for the society as it helps develop advanced tools that assist educators, researchers and software developers to ensure that plagiarism in assignment and projects submitted are detected to avoid compromising on integrity in their work.

[5]   In their survey entitled, 'Plagiarism detection systems' Bin-Habtoor and Zaher have offered a detailed evaluation of the various systems available and their features. Such techniques and various methods for detecting plagiarism have been studied comprehensively in the given state-of-the- art systems. The identified survey provides various techniques including characterization n-gram, vector based, syntactical, semantical, fuzzy, syntacto-semantic, structural-stylometric and cross-linguistic. Through the analysis of these various methods, the research will seek to provide valuable information in regards to the existing plagiarism detection systems and the extent to which, these systems are efficient with an evaluation of the various linguistic patterns and features that may signify cases of plagiarism. By summarizing the present understandings and developments regarding those various systems applied in plagiarism detection, this survey enriches the knowledge of computer theory and engineering on computer science, while emphasizing the evolution of detection methodologies and the integration of new methods. The present study describes the linguistic patterns, textual features and the detection methods employed in these systems providing a detailed picture about how the plagiarism fight is accomplished efficiently. The research will be of significance to faculty and students, as well as scholars and experts in the fields of academic fraud, plagiarism, and information retrieval, to gain insight into the issues surrounding plagiarism detection and the techniques for identifying the existence of plagiarism.

[6]   The study by Sutoyo et al. (2017) examines the detection of document plagiarism through the utilization of the Winnowing algorithm and the k-gram method. The research investigates the effectiveness of these techniques in identifying instances of plagiarism in documents. By combining the Winnowing algorithm with the k-gram method, the study aims to improve the accuracy and efficiency of detecting similarities between texts and uncovering potential cases of plagiarism. This research makes a valuable contribution to the field of artificial intelligence and plagiarism detection by evaluating the synergistic effects of these two methods in detecting document plagiarism.

[7]   The study by Jēkabsons (2020) examines the evaluation of fingerprint selection algorithms for local text reuse detection. The research specifically focuses on assessing and comparing the effectiveness of various fingerprint selection algorithms for the source retrieval stage of local text reuse detection. By analyzing and comparing these algorithms, the study aims to enhance the accuracy and efficiency of detecting instances of text reuse, contributing to the advancement of plagiarism detection methodologies. This research offers valuable insights into the selection of appropriate fingerprinting algorithms for detecting local text reuse, thereby improving the overall effectiveness of text similarity detection systems.

[8]   Zhou et al. (2009) 'Testing Plagiarism in Student Assignment using Winnowing Algorithm', has examined on the effects of Winnowing algorithm in detecting plagiarism on students' assignments. The study was intended to harness the technology of Winnowing algorithm in identifying cases of cheating in student tasks, particularly on programming assignments. As the objective of the study was to determine whether the algorithm could be useful for identifying similarities and possible instances of plagiarism when applied to student assignments, the following research question was used in the study. In thus identifying plagiarism in student submissions, this study adds to the body of knowledge in applied computer science and technology especially with regards to the Winnowing algorithm's efficacy in flagging cases of plagiarism therefore making it very valuable to educational institutions for ensuring academic integrity and originality.

[9]   Online plagiarism checking is another significant field of study that entails the application of spam filters, algorithms, and heuristics in identifying cases of plagiarism. Fahma Rahmatulloh, Listyorini Ahnaf, M. Abdul Kadir, Muhammad Nurul Hidayat, and Margareta Tri Widyawati of Universitas Multimedia Yogyakarta, Indonesia, and 2023 conducted a study to determine the performance of two algorithms, namely Nazief-Adriani and Stemmer Porter, in evaluate plagiarism. Such algorithms were used to investigate the effect of these algorithms in the scenario of Winnowing algorithms on the Indonesian language for plagiarisms identification. This study also showed that the use of the stemming algorithm by Nursalim et al, (2005) specifically the Winnowing algorithm of the language by Nazief -Adriani,

reduced the similarity level but enhanced processing time. Alzahrani et al. (2012) provides an insight into the various techniques applied in detecting important context values, among them being character n-gram, vector based, syntax based, semantic based, fuzzy based, structural based, stylomatic based, and cross-lingual based. In their work, they outlined the need to consider some features of the linguistic patterns and textual elements as some of the key factors that can help in the enhancement of the use of the detection algorithms. Furthermore, Botto-Tobar et al. (2022) also explained the shortcomings of current plagiarism identification methods that are mainly based on extrinsic cleansing methods of identifiable text passages. More advanced techniques and mechanism for detection of plagiarism other than copying checked by this tool were also discussed by them, they also stressed the need of handling following issues in case of cross language plagiarism detection. To conclude, the continued expansion of research in plagiarism detection manifested a focus on self-comparison of the implemented algorithms, studying the language model, and creating more complex methods to combat plagiarism effectively.

[10]   According to Vrbanec & Meštrović (2017), the study on academic plagiarism entailed difficulties in understanding the subject area; especially when tackling methods based on semantic similarity. In order to address this challenge, the study pointed out that other Citation-based Plagiarism Detection developed since plagiarists are known to use portions with citations copied from source documents where or with minimal changes and modifications exercises little control over their contextualization. They include citation processes within scientific documents as complex, language neutral signatures to determine semantic similarity efficiently to detect disguised plagiarism. Gullifer & Tyson (2010) in their empirical research venture undertaking a focus group study on the perception that university students have for the subject of plagiarism. Thus, the insights of students and their knowledge about plagiarism violation can be significant to combat its use effectively. Moreover, in the study conducted by Meuschke et al. (2017), the authors also pointed out that when searching for academic plagiarism, the concept of semantic pattern could have been more focused on. Due to this, the implementation of this solution in plagiarism detection that encompasses different types of similarity features, enhance a broader evaluation of diverse types of plagiarism in academic documents. Thus, the problem of determining ways to combat plagiarism in academic works is most effective taking into account semantic similarity, learners' perception and the use of sophisticated means to address the issue in support of academic integrity.

[11] Nurdiansyah et al. reported a study regarding the Winnowing Algorithm based K-Gram to address the problem of plagiarism particularly in the text-based file document. More specifically, this research has been centered around the employment of K-Gram system within the Winnowing Algorithm to enhance the original anti-plagiarism strategy. K-Gram for example involve the division of the text in to K character chunks to help the algorithm to identify similarities between documents with accuracy. The study further underlined the significance of this approach in improving the identification of plagiarism in textual content matter. In another study, Li et al. looked at the role of machine learning in particularly, in the detection of cases of plagiarism. Their studies specifically focused on the proposed use of machine learning techniques, which include the use of SVM and Neural Networks on textural features and patterns in order to detect plagiarized content. Engaging machine learning to augment the creation of plagiarism detection outcomes that are efficient for tackling big data analysis. Further, Santos et al. are the pioneers who discovered that using data mining and natural language processing (NLP) techniques can also identify plagiarism. To achieve this, they outlined a study to show how NLP techniques including text preprocessing, feature extraction, and similarity measurements could be employed in enhancing the specific outcomes of plagiarism detection systems. As a result of this, machine learning and NLP techniques can help researchers in improving the performance of plagiarism, detection algorithms in identifying affected sections of documents. Altogether, much work is still in progress to develop methods for detecting plagiarism with unique algorithms, machine learning methods, and NLP approaches therefore minimizing the cases of previous and present plagiarism in textual documents.

[12]   In another study, Leman et al. a study was undertaken with the aim of comparing the efficiency of the Rabin Karp and Winnowing algorithms in a statistical analysis of the detection of plagiarism in text documents. These algorithms were used in the research work to address the task of finding similarity in text documents effectively. The use of hash String matching: The Rabin Karp's algorithm which is used for the detection of strings and the Winnowing algorithm for detecting similarity or resemblance k – grammage was used as an aid in the detection of other form of plagiarism textual content. In the similar study, Siahaan et al, proposed the optimization of the Levenshtein distance where it was combined with the Rabin-Karp algorithm in an intense endeavor to enhance the levels of document equivalence. The Rabin-Karp algorithm was outlined and discussed in terms of plagiarism checker because of its capability to detect duplicate figures. Additionally, Mentari et al. assessed and developed a Cross-Language Text Document Plagiarism Detection System using the Winnowing method. Their study focused on the design of a system to detect plagiarism within text documents written in various languages using the methodology of the Winnowing algorithm. The approach exemplifies the effectiveness of the Winnowing algorithm, which is able to detect copied text no matter the language used. Finally, based on

increasing advancements in algorithms such as Rabin Karp and Winnowing, it can be suggested that the research and development of plagiarism detectors and detection methods are still ongoing, which offer considerable scope for detection by comparing and analyzing different text documents to identify similarities.

[13] The study by Ulinnuha et al. (2018) focuses on the implementation of the Winnowing Algorithm for detecting document plagiarism. This research contributes to the field of plagiarism detection by utilizing the Winnowing Algorithm to identify similarities in text documents effectively. Additionally, the study by Duan et al. (2017) discusses a plagiarism detection algorithm based on an extended version of the Winnowing Algorithm, emphasizing the importance of advanced algorithms in detecting instances of plagiarism. Furthermore, Mentari et al. (2022) explore the application of the Winnowing Method for cross-language text document plagiarism detection, showcasing the versatility of this algorithm in identifying copied content across different languages. These studies collectively highlight the significance of utilizing sophisticated algorithms like Winnowing for accurate and efficient plagiarism detection in textual documents.

[14] Gipp et al. (2011) investigated the text- and citation-based comparative methods of identification of plagiarism through GuttenPlag. Hence, their research contained a new method referred to as Citation-based Plagiarism Detection that was tested using a doctoral thesis. The volunteer web-based crowd-sourcing project, GuttenPlag, captured lattice, highly sophisticated cases of copy-paste clearly indicating that despite meticulous inspection, citation-based detection methods can far out perform the other. In another study, Zhou et al. (2020) have proposed a multilevel text alignment approach with cross-document attention where the model outperforms other models for predicting document-to-document relations and sentence-to-document or document-to-sentence relations in tasks such as citation recommendation and plagiarism detection. It compared favorably with traditional hierarchical attention encoders that use recurrent and transformer-based contextualization; it encapsulated the notion of structural alignment across documents as crucial for improving their detection rates. In addition, Alzahrani et al. (2012) presented an excellent survey work involving a discussion on characterization of the major methods for plagiarism detection on character n-gram-based, vector-based, syntax-based, semantic-based, fuzzy-based, structural-based, stylometric-based, and cross-lingual. Analysis of different linguistic patterns and certain features of the text is also important as they can help to improve the effectiveness of the plagiarism detection procedures and help to deal with cases of constructing almost identical texts. Ultimately, striving for an enhancement of efficiency in the identification of plagiarized textual documents,

there is competition in the enhancement of new approaches such as the current citation-based detection, multilevel text alignment detection and comprehensive detection of linguistic patterns.

[15] Meuschke and Gipp Meuschke et al. (2018) conducted a study on the effectiveness of citation pattern matching algorithms for citation-based plagiarism detection. Their research highlighted the importance of combining the analysis of citation patterns and text similarity to improve the identification of concealed academic plagiarism. This approach underscores the significance of leveraging citation patterns in addition to textual content to enhance the accuracy of plagiarism detection algorithms. In a related study, Pertile et al. (2015) developed CitePlag 29, a plagiarism detection system that relies solely on citation patterns and references. This system illustrates the potential of citation-based analysis in effectively detecting instances of plagiarism based on the unique patterns present in citations. Furthermore, Meuschke et al. (2019) focused on enhancing academic plagiarism detection for STEM documents by analyzing mathematical content and citations. Their research emphasized the importance of incorporating mathematical content analysis alongside citation examination to improve the detection of plagiarism in scientific and technical documents.

[16] Al-Suhaiqi et al. (2019) examined the performance of key phrase extraction and Machine Learning technique used for the cross-lingual Arabic-English plagiarism detection in text. In conclusion of the study, the results showed that the Winnowing algorithm was efficient in identifying cross language piracy with high possibility and fewer chances of error from both the sample set and the larger set. Consequently, the present research applies to the area of plagiarism detection by dedicating an investigation towards cross-lingual text reuse detection. Aljohani & Mohd (2014), however, tried an Arabic-English cross-Language Winnowing algorithm to show the importance of such approach to insight cross-Language plagiarism detection cases. In addition, Maqbool et al. (2022) proposed a plagiarism detection system that uses cross-lingual document pairs, thereby emphasizing on feature extraction and the integration of cross-language trained Machine Learning models for better detection of replicated text. The recall by the Winnowing algorithm was 81% as noted in the study by Al-Suhaiqi et al. (2019) .

[17] The study by Nasien & Yansen (2022) focuses on the implementation of the Winnowing Algorithm for detecting plagiarism in the title and abstract of students' final projects. The researchers aimed to establish a website design that checks the similarity of titles and abstracts using the Winnowing method to thoroughly examine the similarity of titles and abstracts in students' final assignments. This research contributes to the field of plagiarism detection by specifically targeting the title and

abstract sections of student projects, showcasing the importance of utilizing advanced algorithms like Winnowing for accurate detection of similarities in academic documents.

[18]   Alzahrani et al. (2011) introduced the "iPlag: Concerning with the research of "Intelligent Plagiarism Reasoner in scientific publications" for the first time in the year 2011. Thus, this system was intended to minimize the role of a human factor and to increase the use of various automated tools for plagiarism identification while utilizing multiple analytical processes. In order the enhance the speed and efficacy of plagiarism detection utilized in scientific articles, iPlag was developed, based on the application of complex analytical algorithms. As a contribution to the research in the domain of plagiarism detection, this work presents an intelligent system to detect plagiarism within documents deposited in the scientific world, therefore contributing to the fight against scientific fraud and maintaining the credibility of information generated in an academic environment.

[19]   Kurniati et al. (2019) also undertook a study where web scraping and Winnowing algorithms were used on the final project titles for assessment for plagiarism. This study set the objective to engage and harness Google Scholar to match the data for final project titles and the adopted methodology was the Winnowing Algorithm. This paper aimed at improving the identification of similarities in the final project title by using the sophisticated employment of algorithmic formulating not only the significance of modern technology in identifying plagiarized content but also the significance of the use of proficiency in identifying plagiarism from academic documents. Other similar studies have not shared the data and utilized Google Scholar sources to compare data for final project titles using Winnowing Algorithm (Kurniati et al. , 2019).

[20] In this work we examine the effectiveness of a dictionary and the employment of n-gram alignment methodologies in refining the cross-language plagiarism detection. The paper focuses on studying the impact of the integration of dictionary and n-gram alignment approaches to enhance CL-plagiarism detection capabilities. In detail, this work examines the improvement of cross-language plagiarism detection at the sub-section level through the utilization of both dictionaries and n-gram alignment techniques. The paper also describes an approach involving the process of dictionary as well as n-gram alignment to enhance CFL-IP's fine-grained accuracy. This paper outlines the methods of using both dictionary and n-gram alignment techniques in improving cross-language anti-plagiarism techniques with increased precision on the fine-grained level.

[21]   Extended Winnowing algorithm has been proposed in 2017 that the authors attempted to improve the flexibility and use of Winnowing algorithm for plagiarism detection (Duan et al. , 2017). Winnowing is a similar algorithm that this algorithm takes cues from and modifies in order to create a more all-encompassing and efficient algorithm that is capable of detecting plagiarism incidences in a range of document types due to its light weight, high efficiency, reliability and flexibility.

[22] The paper by titled "Implementation of Winnowing Algorithm with Dictionary English-Indonesia Technique to Detect Plagiarism" published in the International Journal of Advanced Computer Science and Applications in 2018 focuses on enhancing plagiarism detection using the Winnowing algorithm with an English-Indonesian dictionary technique. The study aims to improve the accuracy and efficiency of plagiarism detection by incorporating dictionary-based methods and n-gram alignment strategies. By leveraging the Winnowing algorithm in conjunction with the English-Indonesian dictionary technique, the research seeks to enhance fine-grained cross-language plagiarism detection, particularly in the context of English and Indonesian texts. This approach showcases a novel method for detecting similarities and potential instances of plagiarism across different languages, contributing to the advancement of plagiarism detection methodologies. The integration of the Winnowing algorithm with the English-Indonesian dictionary technique offers a unique approach to identifying similarities and detecting plagiarism in textual content. By utilizing dictionary-based methods, the study aims to improve the precision and reliability of plagiarism detection systems, especially in scenarios involving cross-language text comparisons. The research by underscores the importance of leveraging linguistic resources and alignment techniques to enhance the performance of plagiarism detection algorithms, showcasing the potential for more accurate and effective detection of plagiarized content. The study by . aligns with the broader research landscape focused on advancing plagiarism detection methodologies through innovative techniques and algorithmic enhancements. By exploring the implementation of the Winnowing algorithm with dictionary-based approaches, the research contributes to the ongoing efforts to improve the accuracy and robustness of plagiarism detection systems. The utilization of the English-Indonesian dictionary technique in conjunction with the Winnowing algorithm demonstrates a tailored approach to cross-language plagiarism detection, catering to the specific linguistic nuances of English and Indonesian texts. In conclusion, the paper by Yudhana et al. presents a novel methodology for detecting plagiarism by combining the Winnowing algorithm with the English-Indonesian dictionary technique. This approach holds promise for enhancing fine-grained cross-language plagiarism detection, offering a valuable

contribution to the field of computer science and advancing the capabilities of plagiarism detection systems in multilingual contexts. The study underscores the significance of leveraging linguistic resources and innovative techniques to improve the accuracy and efficiency of plagiarism detection algorithms, paving the way for more effective identification of plagiarized content in diverse textual datasets.

[23] The paper by titled "Cross-Language Text Document Plagiarism Detection System Using Winnowing Method" published in the Journal of Applied Intelligent System in 2022 focuses on the development of a system for detecting plagiarism in text documents across different languages utilizing the Winnowing method. The study aims to address the challenges of cross-language plagiarism detection by implementing the Winnowing algorithm, a fingerprint-based text similarity detection method. By leveraging the Winnowing method, the research aims to enhance the accuracy and efficiency of detecting plagiarized content in text documents written in diverse languages. This approach showcases a novel method for identifying similarities and potential instances of plagiarism in a cross-language context, contributing to the advancement of plagiarism detection methodologies. The integration of the Winnowing algorithm in the cross-language text document plagiarism detection system offers a promising approach to improving the precision and reliability of plagiarism detection across different languages. By utilizing the Winnowing method, the study aims to enhance the performance of the system in identifying similarities and potential instances of plagiarism in text documents written in various languages. The research by underscores the importance of leveraging advanced algorithms and techniques to improve the accuracy and effectiveness of plagiarism detection systems, particularly in multilingual settings. The study aligns with the broader research landscape focused on advancing plagiarism detection methodologies through innovative approaches and algorithmic enhancements. By exploring the implementation of the Winnowing algorithm in a cross-language context, the research contributes to the ongoing efforts to improve the accuracy and robustness of plagiarism detection systems. The utilization of the Winnowing method in conjunction with cross-language text document analysis demonstrates a tailored approach to detecting similarities and potential instances of plagiarism in diverse linguistic datasets, paving the way for more effective identification of plagiarized content in multilingual environments. In conclusion, the paper by . presents a significant contribution to the field of plagiarism detection by introducing a system that utilizes the Winnowing method for cross-language text document plagiarism detection. This approach holds promise for enhancing the accuracy and efficiency of detecting plagiarism in text documents across different languages, offering valuable insights for researchers, practitioners,

and developers working in the domain of plagiarism detection and text analysis in multilingual contexts. The study highlights the importance of leveraging advanced algorithms and methodologies to address the challenges of cross-language plagiarism detection and improve the overall effectiveness of plagiarism detection systems.

[24]   Kakkonen & Mozgovoy (2010) assessed the current state of hermetic and web plagiarism detection systems for student essays. Their study revealed that many existing plagiarism detection systems lack robust natural language processing (NLP) capabilities, limiting their ability to identify complex forms of plagiarism beyond direct copying or basic editorial alterations. Kakkonen and Mozgovoy integrated NLP techniques into their systems, enabling them to identify more sophisticated forms of plagiarism. Their evaluation underscored the significance of advanced technologies, like NLP, in improving the efficacy of plagiarism detection systems, especially in identifying intricate instances of academic misconduct in student essays.

[25] The paper by titled "The Implementation of Winnowing Algorithm for Plagiarism Detection in Moodle-based E-learning" explores the utilization of the Winnowing algorithm for detecting plagiarism within Moodle-based e-learning environments. The study focuses on enhancing plagiarism detection capabilities within the Moodle platform, a widely used learning management system. By implementing the Winnowing algorithm, the research aims to improve the accuracy and efficiency of identifying plagiarized content in educational settings. This approach offers a tailored solution for detecting plagiarism specifically within the context of e-learning platforms like Moodle, contributing to the advancement of plagiarism detection methodologies in online education environments. 's study highlights the importance of leveraging advanced algorithms like Winnowing for detecting plagiarism in e-learning systems. By integrating the algorithm into Moodle-based platforms, the research aims to provide educators and administrators with a reliable tool for identifying instances of academic dishonesty. The implementation of the Winnowing algorithm within Moodle-based e-learning environments showcases a proactive approach to maintaining academic integrity and upholding ethical standards in online education. The research by . aligns with the broader efforts to enhance plagiarism detection mechanisms in digital learning environments. By focusing on the implementation of the Winnowing algorithm within Moodle, the study addresses the specific challenges associated with detecting plagiarism in online educational platforms. The utilization of advanced algorithms for plagiarism detection in e-learning systems underscores the importance of ensuring academic honesty and originality in digital learning contexts. In conclusion, the paper by Hasan et al. presents a significant contribution to

the field of e-learning by introducing the implementation of the Winnowing algorithm for plagiarism detection within Moodle-based platforms. This approach offers a tailored solution for identifying and addressing instances of plagiarism in online educational settings, emphasizing the importance of academic integrity in digital learning environments. The study by Hasan et al. underscores the value of leveraging advanced algorithms for plagiarism detection to promote ethical conduct and uphold educational standards in e-learning platforms like Moodle.

3. CONCLUSIONS

In conclusion, the survey paper on plagiarism detection using the Winnowing algorithm provides a comprehensive overview of the current state-of-the-art in plagiarism detection systems, with a special emphasis on the Winnowing algorithm and its application in detecting plagiarism. The survey highlights the effectiveness of the Winnowing algorithm in identifying similarities in textual documents by breaking them down into k-gram components and generating fingerprints for comparison. The survey emphasizes the importance of utilizing the Winnowing algorithm for its efficiency, reliability, and flexibility in detecting instances of plagiarism in academic documents.

Furthermore, the survey paper underscores the significance of k-parts in the context of the Winnowing algorithm, where documents are segmented into k-gram components to create fingerprints for comparison. This approach allows for a more granular analysis of textual content, enabling the detection of similarities even in cases of paraphrasing or minor alterations. The survey emphasizes the role of k-parts in enhancing the precision and accuracy of plagiarism detection systems, especially when dealing with subtle forms of academic misconduct.

Moreover, the survey paper highlights the need for continuous advancements in plagiarism detection techniques, including the integration of machine learning, natural language processing, and cross-language approaches to improve the efficacy of detection systems. By leveraging innovative algorithms like Winnowing and emphasizing the importance of k-parts in the detection process, researchers and practitioners can enhance the capabilities of plagiarism detection systems and ensure the integrity of academic work.

In summary, the survey paper provides valuable insights into the application of the Winnowing algorithm and k-parts in plagiarism detection, showcasing their effectiveness in identifying instances of academic misconduct. By focusing on these advanced techniques and methodologies, researchers can further improve the accuracy, efficiency, and reliability of plagiarism detection systems, ultimately contributing to the preservation of

academic integrity and the prevention of plagiarism in scholarly work.

REFERENCES

[1] Aljohani and Mohd "Arabic-English Cross-language Plagiarism Detection using Winnowing Algorithm" Information technology journal (2014)

[2] Velásquez and Taylor "Tools for External Plagiarism Detection in DOCODE" (2014)

[3] Nugroho "Pengaruh Algoritma Stemming Nazief-Adriani Terhadap Kinerja Algoritma Winnowing Untuk Mendeteksi Plagiarisme Bahasa Indonesia" Ultima computing jurnal sistem komputer (2017)

[4] Bandara and Wijayarathna "A Machine Learning Based Tool for Source Code Plagiarism Detection" International journal of machine learning and computing (2011)

[5] Bin-Habtoor and Zaher "A Survey on Plagiarism Detection Systems" International journal of computer theory and engineering (2012)

[6] Sutoyo et al. "Detecting documents plagiarism using winnowing algorithm and k-gram method" (2017)

[7] Jēkabsons "Evaluation of Fingerprint Selection Algorithms for Local Text Reuse Detection" Applied computer systems (2020)

[8] Ramli et al. "Uji Plagiarism pada Tugas Mahasiswa Menggunakan Algoritma Winnowing" Journal of applied computer science and technology (2021)

[9] Rahmatulloh et al. "Comparison between the Stemmer Porter Effect and Nazief-Adriani on the Performance of Winnowing Algorithms for Measuring Plagiarism" International journal on advanced science engineering and information technology (2019)

[10] Vrbanec and Meštrović "The struggle with academic plagiarism: Approaches based on semantic similarity" (2017)

[11] Vrbanec and Meštrović "The struggle with academic plagiarism: Approaches based on semantic similarity" (2017)

[12] Leman et al. "Rabin Karp And Winnowing Algorithm For Statistics Of Text Document Plagiarism Detection" (2019)

[13] Ulinnuha et al. "Implementation of Winnowing Algorithm for Document Plagiarism Detection" Proceeding of the electrical engineering computer science and informatics (2018)

[14] Gipp et al. "Comparative evaluation of text- and citation-based plagiarism detection approaches using guttenplag" (2011)

[15] Gipp and Meuschke "Citation pattern matching algorithms for citation-based plagiarism detection" (2011)

[16] Al-Suhaiqi et al. "Arabic English Cross-Lingual Plagiarism Detection Based on Keyphrases Extraction, Monolingual and Machine Learning Approach" Asian journal of research in computer science (2019)

[17] Nasien and Yansen "Implementation of the Winnowing Algorithm in Detecting Plagiarism in Title and Abstract of Student's Final Project" It journal research and development (2022)

[18] Alzahrani et al. "iPlag: Intelligent Plagiarism Reasoner in scientific publications" (2011)

[19] Kurniati et al. "Web Scraping and Winnowing Algorithms for Plagiarism Detection of Final Project Titles" Lontar komputer jurnal ilmiah teknologi informasi (2019)

[20] Ehsan et al. titled "Using a Dictionary and n-gram Alignment to Improve Fine-grained Cross-Language Plagiarism Detection" (2016)

[21] Duan et al. "A Plagiarism Detection Algorithm based on Extended Winnowing" Matec web of conferences (2017)

[22] Yudhana et al. "Implementation of Winnowing Algorithm with Dictionary English-Indonesia Technique to Detect Plagiarism" International journal of advanced computer science and applications (2018)

[23] Mentari et al. "Cross-Language Text Document Plagiarism Detection System Using Winnowing Method" Journal of applied intelligent system (2022)

[24] Kakkonen and Mozgovoy "Hermetic and Web Plagiarism Detection Systems for Student Essays—An Evaluation of the State-of-the-Art" Journal of educational computing research (2010)

[25] Hasan et al. "The Implementation of Winnowing Algorithm for Plagiarism Detection in Moodle-based E-learning" (2018)