International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Review on Plagiarism Detection Systems, Algorithms, Weakness Points

Khaled Omar[1], Nour Esmaeel[2], ZoAlfekar Ebrahim[3]

Submitted: 25/12/2023 **Revised**: 05/02/2024 **Accepted**: 13/02/2024

Abstract*:* Plagiarism detection is the process of identifying instances of plagiarism in written or digital content. Plagiarism is the act of presenting someone else's work or ideas as one's own without proper attribution or permission, Plagiarism detection can be carried out using various techniques, including manual methods, such as reading and comparing texts, as well as automated methods, such as software-based systems, Automated plagiarism detection systems use algorithms to analyze the content of a document and compare it to other documents in their database or on the internet. Plagiarism detection is important field of natural language processing , artificial intelligence fields, many algorithms and systems have been developed to detect plagiarism, in this paper we will talk about plagiarism detection systems, plagiarism types, plagiarism detection algorithms types , and weakness point of plagiarism detection algorithms , we will talk in details about the most popular plagiarisms detection algorithms which contains string based algorithms, fingerprints based algorithms, semantic based algorithms, syntax based algorithms, deep learning based algorithms, and cross language plagiarism detection algorithms.

Keywords: *Plagiarism Detection Systems, Plagiarism Detection Algorithms.*

1. Introduction

Plagiarism detection is important in the academic society because of its negative impact on the scientific production in academic universities and institutes.

The free content which available on the internet and increasing day by day and this in turn leads to increasing of copy and paste behaviors, and this in turn increase the existence of plagiarism detection and reduction algorithms and systems.

Many systems and algorithms have been developed for plagiarism detection and reduction, and these algorithms use different workflows and techniques. And in this article we will try to cover all details about these systems and algorithms.

In this article we will discuss about plagiarism types, plagiarism detection reducing techniques, plagiarism detection system classifications and structures, plagiarism detection algorithms classifications and workflows, plagiarism detection used technology, plagiarism detection algorithms weakness points, and then the conclusion

2. Plagiarism Classification

Mainly there are two types of plagiarism [1] [2]:

- Source code plagiarism
- Natural language plagiarism

Plagiarism has many classes and these classes different in the behavior of copy and paste from the origin text to suspected one, we will describe in the following the most important types of plagiarism:

Complete plagiarism [3]

This type is the most popular one of plagiarism, in this type of plagiarism someone takes someone works and re-submit it as its origin work without any citation to the source.

Source based plagiarism [4]

This type of plagiarism occurs when someone take someone scientific work and write a wrong reference to the origin work, or write nonexistence reference.

Direct plagiarism [5]

This type of plagiarism occurs when someone copy from another author work word by word without any reference to the source, it is like complete plagiarism, but it refers to sections (rather than all) of another paper.

Paraphrasing plagiarism [6]

This type of plagiarism when someone copy from another author work and change some words and some grammar rules to hide the plagiarism behavior , but the main meaning remain unchangeable.

Self or auto plagiarism [7]

This type of plagiarism occurs when someone uses his previous work and research without any reference to this previous work.

Accidental Plagiarism [8]

Student in universities may do this type of plagiarism without know that this behavior is unacceptable and refused.

Cross language plagiarism [9]

This type of plagiarisms occurs when someone translate another author work to different language and uses it as his/her origin work without any reference to the origin work in the main language.

There is many another classifications of plagiarism, but we mentioned the most popular ones.

[1] *Head of Artificail Intellegence Department, Information Technology Engenerring College, Damascus University,*
kh.om.mail@gmail.com
[2] *Ph.D Student, Ph.D of Web Science Program, Syrian Virtual University*
nour_1263@svuonline.org
[3] *Ph.D Student, Ph.D of Web Science Program, Syrian Virtual University*
alfekar_458@svuonline.org

3. Plagiarisms Systems

In this section we will discuss about plagiarism detection systems classifications, that there are many classifications based on system engine, corpus, submission number, type of text that the system examine it, in the following we will mention the classification of plagiarism detection system based on corpus

There are two main classes of plagiarism detection systems based on corpus

External Plagiarism Detection Systems [10] [11]

In this type of systems the suspected file is compared to the corpus which is already exist in the system database, and this corpus is published on the internet, so when the system finds any similarity between the suspected file and any file from the corpus it adds file reference to the final report which contains all references of the plagiarized papers.

Internal Plagiarism Detection Systems [12] [13]

In this type of systems the suspected file is not compared with any files or corpus, the suspected file is analyzed alone and get its attributes, the analyzing is done by using various algorithms, the most common ones are:

- Stylometric based method which try to find the style of writing of the file,
- Syntax based method which uses syntactical features like parts of speech (POS) tags of sentences or phrases to find out plagiarism,
- Grammar Semantics Hybrid based method

4. Plagiarism Algorithms

In this section we will talk about plagiarism detection algorithms that there are many plagiarism detection algorithms has been developed using various technologies.
We will talking in details about developed algorithms by types, that plagiarism detection algorithms could be classified based on the methodology of the algorithm.

1. String based algorithms[14][15][16]

In these types of plagiarism detection algorithms, the origin file and the suspected file text is compared, the comparison is done by splitting the origin and suspected files text to sentences and then the comparison is done on the sentences levels, the complexity of these algorithms about o (n) 2 that n is the number of text sentences, the most famous algorithms of these type is Longest common String LCS which try to find the longest substring between two series of characters, in the following we will show example of how LCS works:
LCS Example: String1: “AGGTAB”, String2: “GXTXAYB”, the longest common string is GTAB.

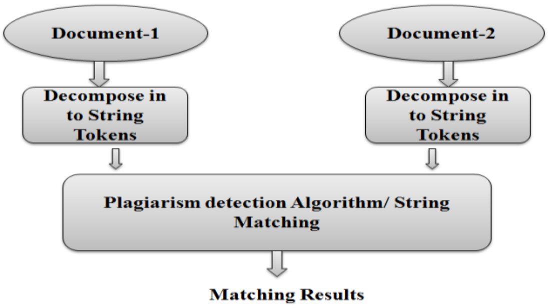
Document-1
Document-2
Decompose in to String Tokens
Decompose in to String Tokens
Plagiarism detection Algorithm/ String Matching
Matching Results

Fig 1: String based plagiarism detection algorithm [16]

2. Fingerprint based algorithms[17] : in these types of plagiarism detection algorithms , the suspected file and the origin file are compared, but the comparison is not done between the files (suspected, origin), is done by comparing files fingerprints, that the main step of these algorithms is firstly generate a fingerprint code for every file , then the comparison is done between the generated fingerprints , the fingerprint could be on the characters level , words level, sentences level, and the generated fingerprint code may be built from the asci codes of the characters or any hashing function, the most popular fingerprint algorithm is Winnowing Algorithm [18] The Winnowing algorithm is utilized for document fingerprint processing, To obtain the hash value of each k-gram, the algorithm is employed to calculate the hash values. The hash rolling function is subsequently used to identify the hash value, and windows are created from these values. In each window, the minimum hash value is chosen, and in the case of multiple hashes with the same minimum value, the rightmost hash value is selected. All of the chosen hash values are then saved to form the document's fingerprint, which serves as the basis for comparing text similarities. The plagiarism detection algorithm employs various methods to ensure that text file matching is not influenced by whitespace, capital letters, punctuation, noise, or word relevance. Matching text files are not dependent on the sequence positions of the words; therefore, even if there are similarities, words with different sequence positions can still be recognized. Winnowing meets these criteria by eliminating irrelevant characters such as punctuation, spaces, and other characters, leaving only letters and numbers for further processing.

Once the Winnowing algorithm generates the fingerprints for each document, it compares the fingerprints of pairs of documents to determine their similarity. The algorithm does this by comparing the set of hash values in each document's fingerprint.
First, the algorithm selects a window size, which is the number of consecutive hash values that are considered at a time. The window slides over the hash values in the fingerprint, and for each window, the algorithm selects the minimum hash value.
The algorithm then creates a set of "minhashes" for each document, which is the set of minimum hash values for all the windows in the document's fingerprint.
To compare two documents, the algorithm calculates their Jaccard similarity [19], which is the size of the intersection of the two documents' sets of minhashes divided by the size of their union. The Jaccard similarity provides a measure of how similar the two documents are based on the hash values in their fingerprints.

$$J(X,Y) = \frac{X \cap Y}{X \cup Y} = \frac{X \cap Y}{|X| + |Y| - |X \cap Y|}$$

Where : X is the first set ,Y is the second set
If the Jaccard similarity between two documents exceeds a certain threshold, the algorithm flags them as potentially plagiarized.
Overall, the Winnowing algorithm provides an effective way to compare documents and detect plagiarism by using fingerprints.

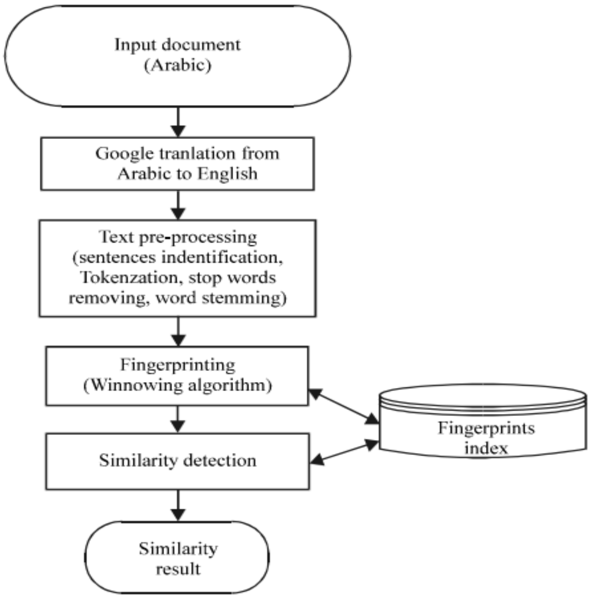

Fig 2: Plagiarism detection using winnowing algorithm [20]

3. Stylometric based algorithms [21][22] : in these types of plagiarism detection algorithms the algorithms try to find a writing features of the author , that a model contains set of writing attributes of the author is built , and then this model is trained basing on the author writing style features, and then when the author submit a new article the model is evaluates this article after analysing it and collecting the submitted author writing style features , and then the model give a decision that the article belongs to the author writing style or not, the writing features attributes are a big set , and various features are used by various developed Stylometric based algorithms, we will mention at the following the most important writing style features :

- Character frequency
- Word length frequency
- Sentence Length Frequency
- Part of Speech Tag Frequency
- Word Specificity Frequency

4. Citation based algorithms[23][24][25]

In these types of plagiarism detection algorithms, the suspected file and the origin file citation patterns are analyzed, and then the generated patterns are compared to detect plagiarism if it exist, the following figure explain the process of citation based algorithms :

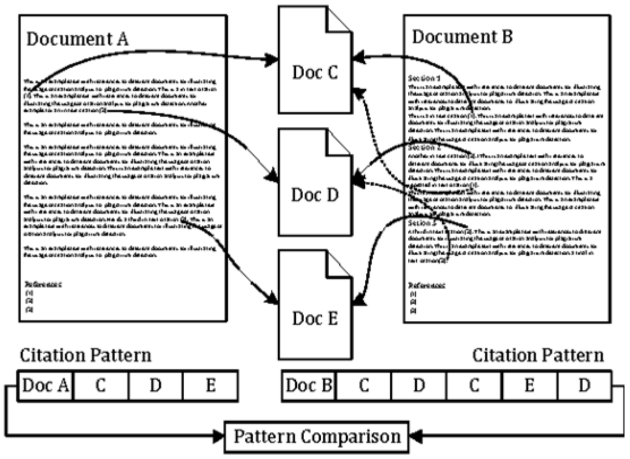

Fig 3: Citation based plagiarism detection algorithms [25]

5. Sematic based algorithms:

In these type of plagiarism detection algorithms the analyzing of suspected text is done semantically, these algorithms may semantic dictionaries or semantic ontologies or both of them,

- Semantic based algorithms based on semantic Dictionaries [26]: The main cause of using of semantic dictionaries is to overcome to the behavior of the plagiarized person , that when someone steal others work he try to replace many words by their synonyms to try to hide the plagiarism action, and by using semantic dictionaries the semantic based plagiarism detection algorithm can detect plagiarism in spite of replacing words by its synonyms, the most popular semantic dictionary is WordNet[27], it supports multi languages , and there is an Arabic version of it called Arabic WordNet, the complexity of these algorithms is o(n)2 , that these algorithms compare every sentence from the suspected file with origin file sentences , and the comparison on the sentences level is done after enrichment the sentences with words synonyms from semantic dictionaries.
- Semantic based algorithms based on semantic ontologies[28]: these types of algorithms use specific domain ontologies to detect plagiarism, such as medical ontologies to detect plagiarism in medical papers, these algorithms benefits from the structure of the ontologies which describe the concepts (definitions, relations, associations, etc.) to analyzing and calculating similarity between suspected file and origin file,

6. Syntax tree based algorithms[29]:

These types of plagiarism detection algorithms are developed to detect plagiarism in programming languages source code,
That these algorithms generate abstract syntax tree for the source code for suspected and origin file, ASTs capture the structural features of code snippets and represent them in a hierarchical tree structure, ASTs are more suitable for code snippet searching because they represent source code at an abstract level where unimportant elements such as grammar symbols can be ignored as noise. Therefore, only the structure and general meaning of source files are considered for comparison, in the following image abstract syntax tree for if statement in code

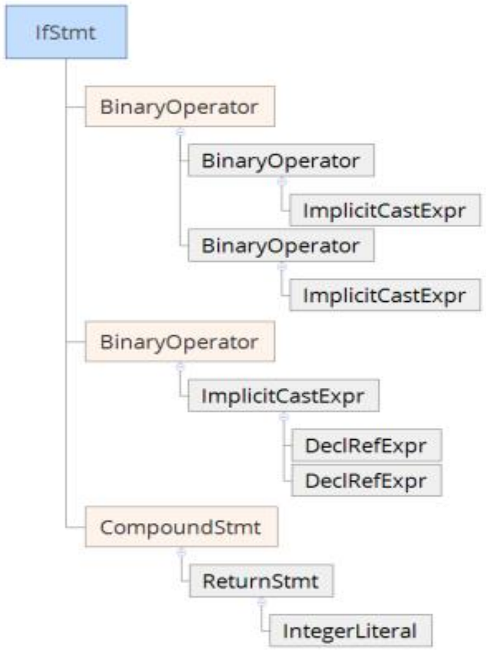

Fig 4: Abstract syntax tree for if-statement [30]

7. Statistical based algorithms[31]:

These types of plagiarism detection algorithms depends on statistical features of texts to detect the suspected files, the most important features that used in such algorithms are the follow:

- Terms frequency (TF)
- Average token length.
- Average sentence length
- Nouns count
- Verbs count
- Prepositions count
- Pronouns count
- Stopwords count
- Number of characters.
- Sentences counts,
- Sentences lengths,
- Grammatical mistakes.

Many algorithms have been developed of this type, in these types of algorithms the similarity is calculated between suspected file and origin file without taking in consideration the order of terms occurrences in files.

8. Deep learning algorithms[32][33]:

In these types of plagiarism detection algorithms, the developed algorithms depends on deep learning and machine learning technology, which are important fields of the artificial intelligence, Deep learning algorithms play a crucial role in computational intelligence, specifically in the machine learning domain. These algorithms utilize multi-layered processing models to learn data representations with varying levels of abstraction. Their success has been demonstrated in various fields, including speech and image recognition, object detection, and NLP. In NLP, different deep learning architectures have been utilized, such as simple Neural Networks for word embedding, as well as more complex algorithms like Siamese LSTM for object similarity detection, Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs), which have proven to be highly effective in classification tasks.

In [33] Authors proposed a plagiarism detection algorithm contains two machine learning classifiers , Siamese LSTM for learning documents similarity , Convolutional neural network CNN classifier to make accrue plagiarism type classification, that the first LSTM classifier takes as input pairs of origin and plagiarized documents ,before this classifier run , a step for documents preprocessing id done which includes stop words removing, stemming , sentences segmentation,doc2vector conversions, after the preprocessing stage is finished the vectors of pairs origin, plagiarized documents are entered to the LSTM classifier and then Each pair's LSTM representation will be labeled by a one hot vector that illustrates a type of plagiarism. And to consolidate proposed approach Authors added another phase which concerns a classification of the types of plagiarism learned in the first part by adding CNN classifier.

9. Word2vector based algorithms[34]:

In these types of algorithms, the suspected and origin files are converted to numbers using word2vector or doc2vector using vector embedding that's metadata for machine learning algorithms to determine similarities between various words, the most popular language model is GloVe[35] which is an unsupervised learning algorithm for obtaining vector representations for words. Training is performed on aggregated global word-word co-occurrence statistics from a corpus, and the resulting representations showcase interesting linear substructures of the word vector space.

After files (suspected file, origin file) represented in word2vector model the stage of detect similarity is done by calculating the similarity between sentences vectors (suspected file sentences vectors, origin file sentence vectors) if the similarity score is greater than a specific threshold the two sentences are matched to be similar and plagiarism is detected between it

10. Hybrid based algorithms:

In these types of algorithms the developed systems are mixture of many algorithms to increase the accuracy and the efficiency of the plagiarism detection systems, in [36] authors proposed a hybrid plagiarism detection algorithm as shown in the following figure:

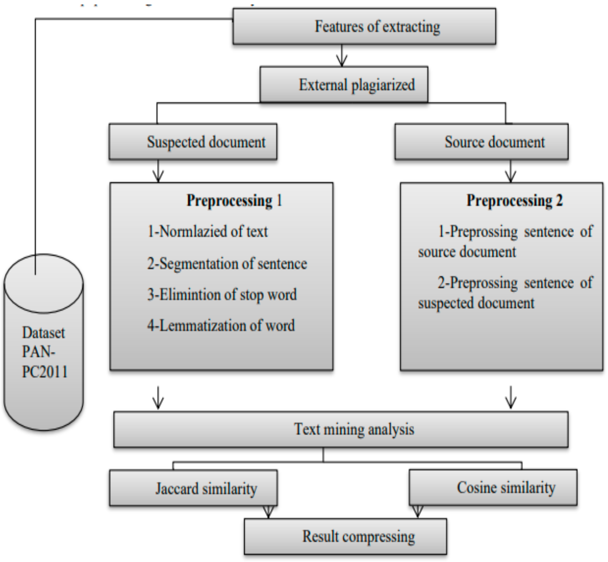

Fig 5: hybrid system for plagiarism detection [36]

The authors have used two similarity calculations, Jaccard similarity, and Cosine similarity.

A hybrid system uses the following relationship that based on the equation of Jaccard and cosine Measurement and the algorithm threshold:

- Jaccard similarity threshold (Tj),
- Cosine similarity threshold (Tc).

$$Plag = \{s: s \in d, \alpha_s \oplus \beta_s = 1\}$$

Where (s) is the suspicious sentence in document D, ⊕ is a logical OR operation

α_s Is Cosine similarity Score (between suspected sentence, origin sentence)

$$\alpha_s = \begin{cases} 0\ if C_s > Tc \\ 1, if C_s < Tc \end{cases}$$

β_s Is Jaccard similarity Score (between suspected sentence, origin sentence)

$$\beta_s = \begin{cases} 0\ if j_s > Tj \\ 1, if j_s < Tj \end{cases}$$

11. Cross Language algorithms

In these algorithms the origin and suspected files are translated to one language (ordinary language of the origin file), and then one plagiarism detection method is applied to detect plagiarized sentences.

In [37] Authors proposed a system for Arabic –English plagiarism detection which translate Arabic files to English then make the text preprocessing step which includes sentences detection, tokenization, stop words removing, word stemming, and then the system try to detect plagiarism using fingerprint algorithm (winnowing algorithm) to generate fingerprint code for suspected and origin file, and then calculates the similarity between generated fingerprints, the following figure shows the proposed system workflow.

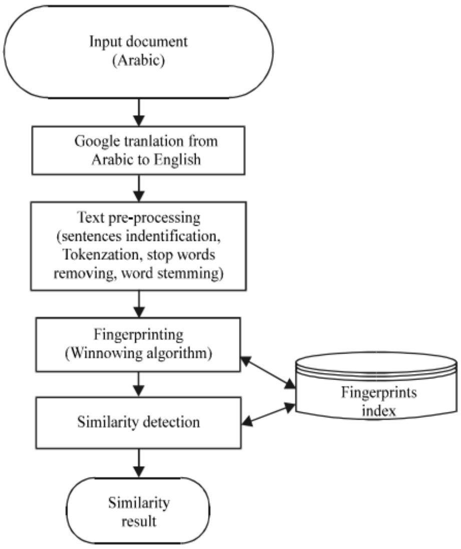
Input document
(Arabic)
Google tranlation from Arabic to English
Text pre-processing (sentences indentification, Tokenzation, stop words removing, word stemming)
Fingerprinting (Winnowing algorithm)
Fingerprints index
Similarity detection
Similarity result

Fig 6: Cross language plagiarism system [37]

5. Plagiarism Algorithms Analyzing

In this section we will talk in details about the weakness points of mentioned plagiarism detection type algorithms.

- String based algorithms: these algorithms highly affected with words order in suspected and origin files, that it cannot detect many plagiarism cases when the suspected file contains plagiarized text with a different words order from the origin file.
- Fingerprint based algorithms: these algorithms complexity is high that it contains a stage for generate fingerprint code for files, and a stage for fingerprint comparisons, and it affected by the words order in suspected file and origin file.
- Stylometric based algorithms: these algorithms try to build a model for specific author , and it needs a corpus of the author to analyzing and building the author model , and then it return a feedback if a paper is belongs to authors articles or not , this means that these types of algorithms are not generic for detect plagiarism to all authors.
- Citation based algorithms: these types of algorithms highly affected by the changes on the references orders and citation false information, that when users copy and paste text and change the citation references information to another ones, in these cases these algorithms will be weak in plagiarism detection.
- Sematic based algorithms: these types of algorithms are effective in common, but for usage these algorithms we always need valued domain ontologies to benefit from the strangeness of these algorithms in plagiarism detection.
- Syntax tree based algorithms: these types of algorithms are strong in source code plagiarism detection, although that these algorithms are used in natural languages plagiarism detection but it have a weakness point generated from the ambiguity in natural languages, that this led to generate more than one abstract syntax tree to the same part of text, this in turns affect the accuracy of these algorithms in plagiarism detection.
- Statistical based algorithms: these types of algorithms depend on statistical texts information of the (suspected, origin) files, so when these features changed or altered in the plagiarized text, in this cases these algorithm will be weak in plagiarism detection.
- Deep learning algorithms: these algorithms are strong and efficient in plagiarism detection, but it needs a training phase for building the model before using it in plagiarism detection.
- Word2vector based algorithms: these algorithms are very strong and efficient in plagiarism detection, but it mainly depend on the word2vector used model and it is limited to the capacity of the used word2vector model.
- Hybrid based algorithms: these types of algorithms used many mixture of algorithm to plagiarism detection, and this is done to get rid of the weakness points of various plagiarism detection algorithms, and it very strong and efficient algorithms.
- Cross Language algorithms: these types of algorithms use a mixture of plagiarism detection types, these algorithms are highly affected by the automatic translation between suspected and origin files.

6. Conclusion

In conclusion, plagiarism is a serious ethical violation that involves presenting someone else's work or ideas as one's own. It undermines the originality and integrity of written work and can have serious consequences, including academic penalties, legal action, damage to reputation, and loss of credibility.

To avoid plagiarism, it is important to understand what it is and how to properly cite and reference sources. This includes giving credit to the original author and properly paraphrasing or summarizing their ideas.

In today's digital age, plagiarism detection technology has become an important tool in identifying instances of plagiarism. However, it is important to note that these tools are not foolproof and should be used in conjunction with manual methods and critical thinking. Ultimately, upholding the principles of academic integrity and ethical writing practices is essential to maintaining a culture of intellectual honesty and promoting the advancement of knowledge and scholarship.

In this paper we have classified the plagiarism detection algorithms in general types, based on how these algorithms work, and then we analyzed these types of algorithms classes and view the weakness point and strengthens ones for each algorithms types.

References

[1] Khan, Nosheena & Agrawal, Chetan & Ansari, Tehreem. (2018). A Review on Various Plagiarism Detection Systems Based on Exterior and Interior Method. IJARCCE. 7. 6-12. 10.17148/
IJARCCE.2018.792.

[2] Citation-based Plagiarism Detection - Detecting Disguised and Cross-language Plagiarism using Citation Pattern Analysis - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Classification-of-Plagiarism-Detection-Approaches_fig2_262689913
[accessed 24 May, 2023]

[3] 7 Common Types of Plagiarism, With Examples. (2022,

June 2). 7 Common Types of Plagiarism, With Examples | Grammarly Blog. https://www.grammarly.com/blog/types-of-plagiarism/

[4] WHAT IS PLAGIARISM – CIBNP. (n.d.). https://www.cibnp.com/ what-is-plagiarism,(2022, June 2)

[5] Direct Plagiarism and How to Avoid it. (2022, July 7). FixGerald.com. https://fixgerald.com/blog/direct-plagiarism

[6] What is the difference between plagiarism and paraphrasing? (n.d.). Enago. https://www.enago.com/plagiarism-checker/resources/ difference-between-plagiarism-and-paraphrasing.htm

[7] Writer. (2023, January 6). Prevent plagiarism before hitting publish - Writer. https://writer.com/guides/plagiarism/

[8] Carmil. (2022). How to Understand and Avoid Accidental Plagiarism Using a Plagiarism Checker? Copyleaks. https://copyleaks.com/blog/accidental-plagiarism-understanding-and-avoiding-it

[9] Tlitova A, Toschev A, Talanov M and Kurnosov V (2020) Meta-Analysis of Cross-Language Plagiarism and Self-Plagiarism Detection Methods for Russian-English Language Pair. Front. Comput. Sci. 2:523053. doi: 10.3389/fcomp.2020.523053

[10] Abdi, A., Shamsuddin, S. M., Idris, N., Alguliev, R. M., & Aliguliyev, R. M. (2017). A linguistic treatment for automatic external plagiarism detection. Knowledge-based systems, 135, 135-146. https://doi.org/10.1016/j.knosys.2017.08.008

[11] Zechner, Mario et al. “External and Intrinsic Plagiarism Detection Using Vector Space Models.” (2009).

[12] Checkforplag. (n.d.). Internal and External Plagiarism. https://www.checkforplag.com/Internal-and-external-plagiarism

[13] J. J. G. Adeva, N. L. Carroll and R. A. Calvo, "Applying Plagiarism Detection to Engineering Education," 2006 7th International Conference on Information Technology Based Higher Education and Training, Ultimo, Australia, 2006, pp. 722-731, doi: 10.1109/ITHET.2006.339692.

[14] W. G. S. Parwita, I. G. A. A. D. Indradewi and I. N. S. W. Wijaya, "String Matching based Plagiarism Detection for Document in Bahasa Indonesia," 2019 5th International Conference on New Media Studies (CONMEDIA), Bali, Indonesia, 2019, pp. 54-58, doi: 10.1109/CONMEDIA46929.2019.8981821.

[15] Pandey, K.L., Agarwal, S., Misra, S., Prasad, R. (2012). Plagiarism Detection in Software Using Efficient String Matching. In: , et al. Computational Science and Its Applications – ICCSA 2012. ICCSA 2012. Lecture Notes in Computer Science, vol 7336. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31128-4_11

[16] Importance of String Matching in Real World Problems - Scientific Figure on ResearchGate. Available from: https://www.researchgate. net/figure/Plagiarism-Detection-System_fig8_304305210 [accessed 20 May, 2023] 26-WordNet, “About WordNet,” http://wordnet. princeton.edu/2010

[17] Narayanan, Sandhya & Surendran, Simi. (2012). Source code plagiarism detection and performance analysis using fingerprint based distance measure method. ICCSE 2012 - Proceedings of 2012 7th International Conference on Computer Science and Education. 1065-1068. 10.1109/ICCSE.2012.6295247.

[18] E. G. Hasan, A. Wicaksana and S. Hansun, "The Implementation of Winnowing Algorithm for Plagiarism Detection in Moodle-based E-learning," 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), Singapore, 2018, pp. 321-325, doi: 10.1109/ICIS.2018.8466429.

[19] Jaccard Similarity. (n.d.). https://www.learndatasci.com/glossary/ jaccard-similarity/

[20] Adel Aljohani and Masnizah Mohd, 2014. Arabic-English Cross-language Plagiarism Detection using Winnowing Algorithm. Information Technology Journal, 13: 2349-2355.

[21] Krause, Markus. (2015). Stylometry-based Fraud and Plagiarism Detection for Learning at Scale.

[22] How to CiteSylvia Putri Gunawan, Lucia Dwi Krisnawati, & Chrismanto, A. R. (2020). Analysis of Stylometric Features and Segmentation Strategies in Intrinsic Plagiarism Detection System. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 4(5), 988-997. https://doi.org/10.29207/resti.v4i5.2486

[23] Gipp, Bela & Meuschke, Norman. (2011). Citation Pattern Matching Algorithms for Citation-based Plagiarism Detection: Greedy Citation Tiling, Citation Chunking and Longest Common Citation Sequence.. DocEng 2011 - Proceedings of the 2011 ACM Symposium on Document Engineering. 249-258. 10.1145/ 2034691.2034741.

[24] Gipp, Bela & Beel, Joeran. (2010). Citation based Plagiarism detection: A new approach to identify plagiarized work language independently. HT'10 - Proceedings of the 21st ACM Conference on Hypertext and Hypermedia. 273-274. 10.1145/1810617.1810671.

[25] Citation-based Plagiarism Detection – Idea, Implementation and Evaluation Gipp, B. (2012)
Bulletin of the IEEE Technical Committee on Digital Libraries, 8(1).

[26] Sharma, Kamlesh & Garg, Nidhi & Pandey, Arun & Yadav, Daksh & Nikhil,. (2021). Plagiarism Detection Technique using www and Wordnet. Indian Journal of Artificial Intelligence and Neural Networking. 1. 1-6. 10.35940/ijainn.B1015.061321.

[27] WordNet, “About WordNet,” http://wordnet.princeton.edu/2010

[28] Shenoy, Manjula. (2012). Semantic Plagiarism Detection System Using Ontology Mapping. Advanced Computing: An International Journal. 3. 59-62. 10.5121/acij.2012.3306.

[29] M. Duracik, P. Hrkut, E. Krsak and S. Toth, "Abstract Syntax Tree Based Source Code Antiplagiarism System for Large Projects Set," in IEEE Access, vol. 8, pp. 175347-175359, 2020, doi: 10.1109/ACCESS.2020.3026422.

[30] Torres, R., Kunkel, J. M., Dolz, M. F. and Ludwig, T. (2018) Comparison of Clang Abstract Syntax Trees using string kernels. In: CADO 2018, 16-20 July, Orleans, France, pp. 106-113. Available at http://centaur.reading.ac.uk/79588/

[31] Bamidis PD, Lithari C, Konstantinidis ST. Revisiting Information Technology tools serving authorship and editorship: a case-guided tutorial to statistical analysis and plagiarism detection. Hippokratia. 2010 Dec;14(Suppl 1):38-48. PMID: 21487489; PMCID: PMC3049420.

[32] El-Rashidy, M.A., Mohamed, R.G., El-Fishawy, N.A. et al. Reliable plagiarism detection system based on deep learning approaches. Neural Comput & Applic 34, 18837–18858 (2022). https://doi.org/10.1007/s00521-022-07486-w

[33] El Mostafa Hambi, Faouzia Benabbou,A New Online Plagiarism Detection System based on Deep Learning,(IJACSA) International Journal of Advanced Computer Science and Applications Vol. 11, No. 9, 2020

[34] K. Omar and A. Hilal, "Plagiarism Detection in Arabic Documents using word2vector and Arabic WordNet," 2022 International Arab Conference on Information Technology (ACIT), Abu Dhabi, United Arab Emirates, 2022, pp. 1-5, doi: 10.1109/ACIT57182. 2022.9994090.

[35] Pennington, J. (n.d.). GloVe: Global Vectors for Word Representation. https://nlp.stanford.edu/projects/glove/

[36] Khiled, Farah & Al-Tamimi, Mohammed. (2021). Hybrid System for Plagiarism Detection on A Scientific Paper. Turkish Journal of Computer and Mathematics Education (TURCOMAT). 12. 5707-5719.

[37] Adel Aljohani and Masnizah Mohd, 2014. Arabic-English Cross-language Plagiarism Detection using Winnowing Algorithm. Information Technology Journal, 13: 2349-2355.