

The Effectiveness of Plagiarism Detection Software in Protecting Intellectual Property: Testing the Efficiency of Turnitin in Detecting Plagiarism in Texts Written in Arabic

Dr. Khaled Baroudi

École Normale Supérieure of Bouzareah – Algeria

Email: khaled.baroudi@ensb.dz

Received : 21/12/2025 ; Accepted : 28/05/2026 ; Published : 17/06/2026

Abstract

The protection of intellectual and literary property has become a major concern for researchers and a challenge to ensuring academic integrity and adherence to research ethics. Among the measures adopted to achieve this objective is the submission of scientific articles uploaded through the ASJP platform to plagiarism detection before they are sent for peer review. This study aims to examine the effectiveness of the Turnitin software in measuring similarity and detecting plagiarism in texts written in Arabic, particularly in light of the flexible nature and distinctive characteristics of the Arabic language, as well as the challenges posed by artificial intelligence.

Keywords: intellectual property; protection; academic plagiarism; software; artificial intelligence.

1. Introduction

Thought is a uniquely human faculty whose use is accessible to everyone, regardless of language, belief, or race. Benefiting from it is a common right, and every individual has a legitimate share in it that no one else may dispute. Appropriating another person's intellectual work without permission or attributing it to someone other than its rightful author constitutes an infringement upon another's property. In this respect, there is no difference between such an act and assaulting someone physically, stealing another person's wealth, or violating another's honor. Although the nature of these rights differs whether material or moral a right remains a right.

Viewing intellectual property as being of lesser importance than tangible property has paved the way for treating violations of material property as crimes while regarding infringements of intellectual property as insignificant incidents. Consequently, the same offense has come to be judged by two different standards, even though the violation itself is fundamentally the same. Worse still, the overwhelming majority of people fail to recognize this type of property, whether out of ignorance or deliberate disregard. This has led us into the quagmire of normalized intellectual theft, only for us to realize, belatedly, the necessity of changing this perception and enacting laws and conventions that protect and preserve this right.

Once the necessity of safeguarding this right became widely acknowledged, it also became imperative to consider practical means of achieving such protection while taking into account technological developments, including the rapid flow and transfer of data and its enormous volume. Under these circumstances, human effort alone may no longer be sufficient. It has therefore become essential to harness technology to provide effective solutions, a role that plagiarism detection and similarity-checking software seeks to fulfill.

1.1 Research Problem

This study addresses the following research problem:

How can intellectual and literary property be protected in light of contemporary challenges through the utilization of artificial intelligence technologies?

1.2 Research Questions

The following questions arise from the main research problem:

What is meant by intellectual and literary property?

What are the challenges facing the protection of researchers' intellectual and literary property?

How can artificial intelligence mechanisms be employed to protect intellectual and literary property from plagiarism?

Does the nature of the Arabic language constitute an obstacle to, or a facilitator of, the implementation of intellectual property protection?

1.3 Research Objectives

The present study seeks to achieve the following objectives:

To demonstrate the importance of protecting intellectual property in advancing scientific research and safeguarding intellectual assets.

To identify the artificial intelligence mechanisms employed to protect researchers' intellectual property.

To assess the effectiveness of plagiarism detection software in identifying cases of academic plagiarism.

1.4 Research Methodology

This study adopted two research approaches: the descriptive and the analytical methods.

Descriptive Method: This method was employed to describe the operating mechanism of the Turnitin similarity detection software.

Analytical Method: This method was adopted to analyze how plagiarism detection software functions and to examine the impact of the structural characteristics of the Arabic language on the ability of such software to detect academic plagiarism.

1.5 Research Terminology

Intellectual Property

Intellectual property is a descriptive compound consisting of two words: property and intellectual.

Property is a verbal noun derived from the triliteral verb *malaka* (to own), meaning possession of something and the exclusive right to dispose of it (Academy of the Arabic Language in Cairo, 2004, p. 886), or the right to use and dispose of a thing without restriction (Ahmed Mukhtar, 2008, p. 2123). The Algerian Civil Code defines it as the right to enjoy and dispose of property, provided that such use does not violate laws and regulations (Algerian Civil Code, Article 674, 2007, p. 110).

Intellectual is a derived noun from the triliteral verb *fakkara* (to think), in the intensified form (*fa''ala*), and refers to everything produced by the human mind, including ideas, expressions, and creative works.

Accordingly, intellectual property refers to the right to enjoy and dispose of intellectual production, in contrast to material property, which encompasses tangible objects perceived through the senses.

Academic Plagiarism

Academic plagiarism is a descriptive compound consisting of two words: plagiarism and academic.

Plagiarism is derived from the triliteral verb saraqa (to steal), meaning the unauthorized and clandestine taking of another person's possessions (Al-Jurjani, 1983, p. 118). Academic plagiarism refers to taking another person's ideas without acknowledging their source or attributing them to someone other than their original author.

Characteristics of the Arabic Language

The Arabic language possesses a number of distinctive features that differentiate it from other natural languages, including characteristics related to its phonology, vocabulary, morphology, and syntax. In the present study, the focus is specifically on its morphological and syntactic characteristics, particularly those governing the arrangement and structure of linguistic expressions.

Similarity Index Measurement

Traditionally, quotation (iqtibās) is defined as incorporating into prose or poetry passages from the Holy Qur'an or the Prophetic Sunnah without explicitly indicating their origin (Al-Tahanawi, 1996, p. 242). Incorporating another person's words, however, was traditionally referred to as talmīḥ (allusion). In contemporary usage, the term quotation has come to denote the incorporation of another author's words while explicitly acknowledging the original source in order to avoid academic plagiarism.

Similarity index measurement is the process of calculating the percentage of borrowed material in a text by determining the ratio of quoted expressions to the total number of expressions in the document.

2. Protection of Intellectual Property

Numerous mechanisms have been developed to protect intellectual property. Among the most important are the following:

2.1 Legislative and Legal Provisions

These include laws, ordinances, and regulations enacted to protect authors' rights from infringement or theft. In Algerian legislation, examples include the following (Official Gazette of the People's Democratic Republic of Algeria, 2003, pp. 3–35):

Ordinance No. 03–05, dated 19 July 2003, relating to copyright and related rights, promulgated on 23 July 2003.

Ordinance No. 03–06, dated 19 July 2003, relating to trademarks, promulgated on 23 July 2003.

Ordinance No. 03–07, dated 19 July 2003, relating to patents, promulgated on 23 July 2003.

2.2 International Conventions

These comprise the internationally recognized standards adopted by member states for the protection of intellectual property. Among the most notable are (Khawadjiya, 2022, pp. 7–8):

The Berne Convention for the Protection of Literary and Artistic Works, signed in 1886.

The Universal Copyright Convention, signed on 6 November 1952.

The Washington Treaty on Intellectual Property in Respect of Integrated Circuits, concluded on 26 May 1989.

2.3 Declaration of Commitment to Academic Integrity

This is a formal document through which a researcher pledges to uphold academic integrity and observe honesty throughout the research process.

2.4 Similarity Detection Software

Similarity detection software consists of electronic applications designed to measure the level of textual similarity in scientific and academic research. These programs contribute to the protection of intellectual property by establishing a maximum permissible similarity threshold that researchers are required not to exceed in their scholarly work.

3. Applied Study

3.1 Research Problem

The applied study investigates the effectiveness of similarity detection software in protecting intellectual property.

3.2 Research Questions

The following questions arise from the research problem:

How can similarity detection software be used as a means of protecting intellectual property?

What is the operating mechanism of these software applications?

Does the nature of the Arabic language positively or negatively affect the ability of these programs to detect plagiarism?

To what extent are these programs effective in detecting academic plagiarism?

3.3 Research Hypotheses

First Hypothesis: Similarity detection software is employed as a means of protecting intellectual property through two fundamental principles: encouraging creativity and innovation by establishing a maximum permissible similarity threshold that researchers must not exceed, and detecting plagiarism and academic misconduct.

Second Hypothesis: These programs measure the similarity index by comparing the structure of expressions (i.e., the sequence of words). Each program applies its own criteria; for example, some consider a sequence of six consecutive matching words as evidence of quotation, whereas others adopt different thresholds that may be higher or lower.

Third Hypothesis: The characteristics of the Arabic language negatively affect the ability of these programs to detect plagiarism because Arabic is highly flexible and allows users to express the same idea through an almost unlimited variety of formulations (paraphrasing).

Fourth Hypothesis: The effectiveness of plagiarism detection software depends on the extent and quality of the scientific materials contained in its database.

3.4 Research Objectives

This study seeks to achieve the following objectives:

To evaluate the effectiveness of similarity detection software.

To determine the impact of the characteristics of the Arabic language on plagiarism detection.

3.5 Research Sample

The study sample consisted of the Turnitin similarity detection software.

3.6 Scope of the Study

The study was conducted during the second semester of the year 2024.

4. Turnitin Software

4.1 Overview of the Software

Turnitin is a similarity detection program designed to measure the similarity index in various forms of scholarly work, including abstracts, journal articles, theses, dissertations, and other academic documents. It is a commercial (paid) software application.

The platform allows users to upload files for similarity analysis in several formats, including Word, PDF, and URL, through a user-friendly interface that supports multiple languages.

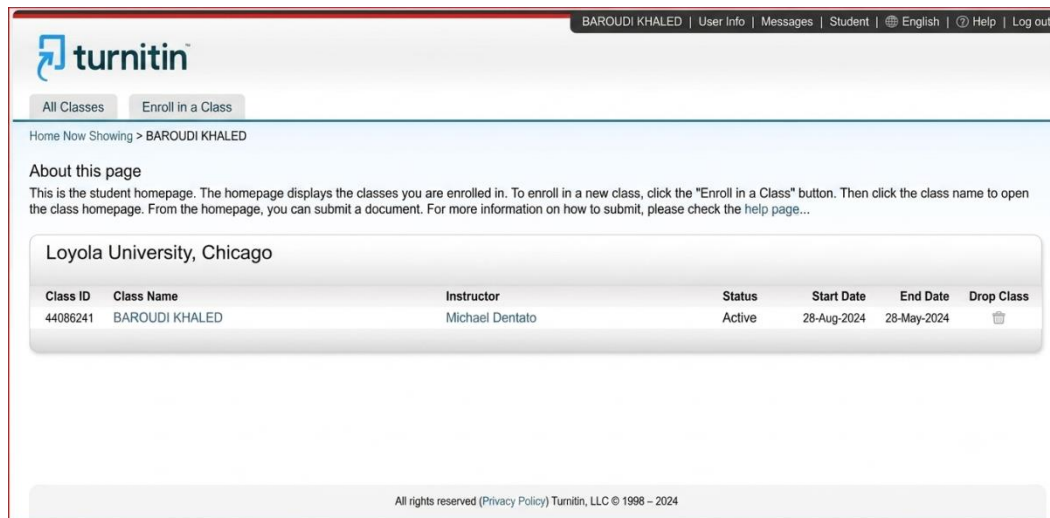


Figure (1): Screenshot of the Turnitin similarity detection interface.

(https://www.turnitin.com/s_home.asp?lang=ar, 2024)

4.2 Operating Mechanism of the Software

The similarity index is measured by the software through the following stages:

4.2.1 Input Stage

At this stage, the text, document, or URL of the work to be analyzed is submitted to the software. Each type of input is uploaded through its designated icon within the user interface.

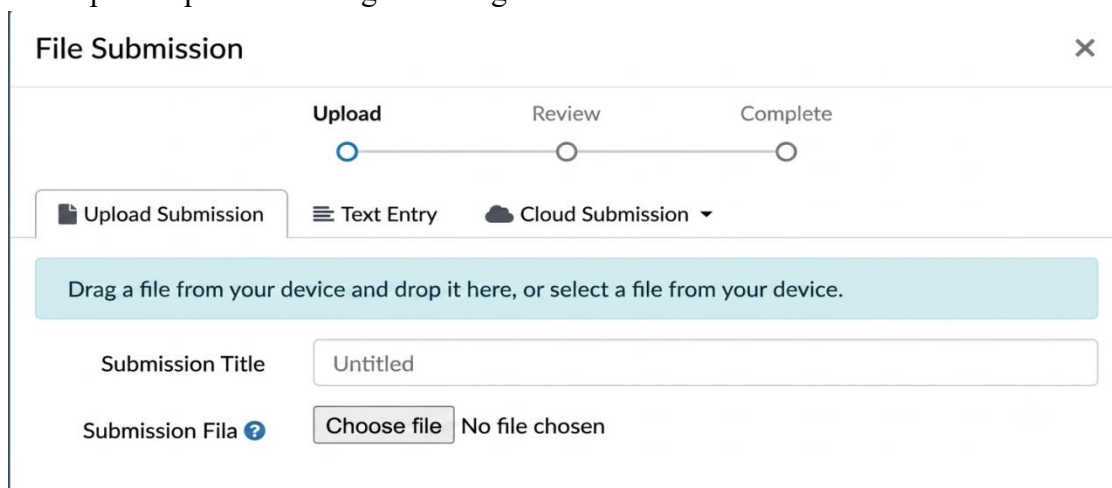


Figure (2): Screenshot illustrating the first stage of similarity measurement in Turnitin.

4.2.2 Similarity Analysis Stage

During this stage, the software compares the sequence of words in the submitted document with the sequences of words contained in the documents stored in its database. Whenever a matching sequence is identified, the software marks the corresponding passage as containing quoted or matching text. Upon completion of the comparison process, it calculates the similarity index relative to the total text using the following formula:

Similarity Index=

Total Number of Words in the Text

Number of Matching Words×100

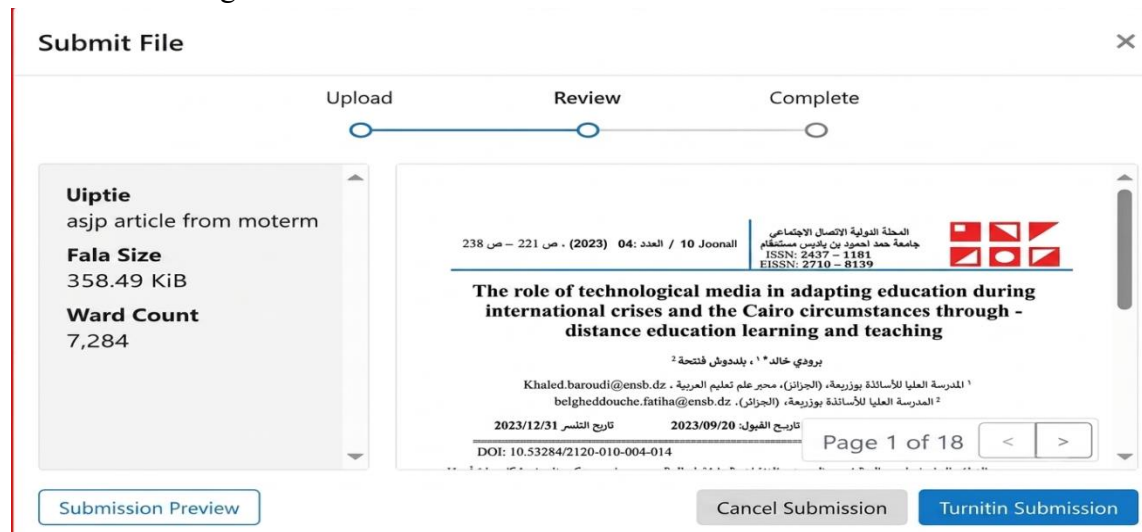


Figure (3): Screenshot illustrating the second stage of similarity measurement in Turnitin.

4.2.3 Results Display Stage

After completing the similarity analysis, the software generates a summary report indicating the overall similarity percentage and the percentage of original (non-matching) text. It also highlights the matched passages using a specific color.

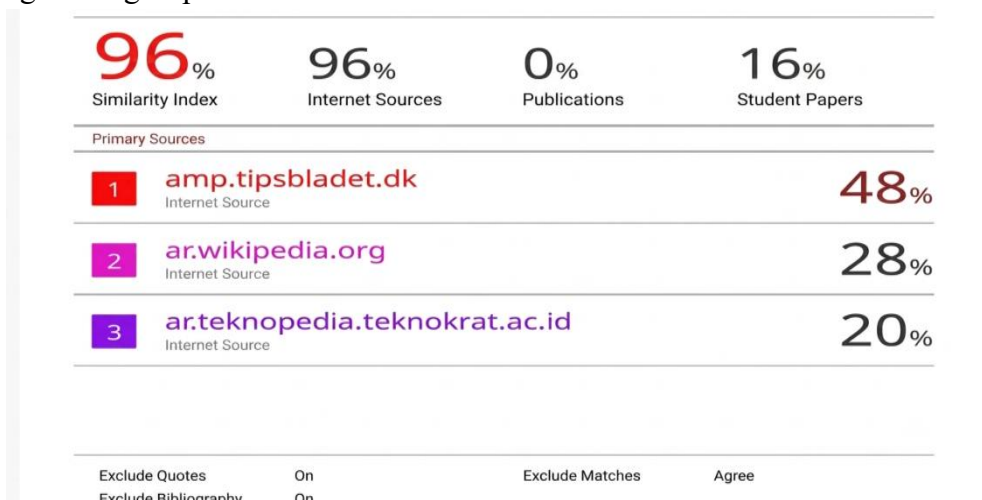


Figure (4): Screenshot illustrating the final stage of similarity measurement in Turnitin.

5. Testing the Efficiency of the Software

5.1 Testing the Software's Efficiency in Detecting Academic Plagiarism

The efficiency of the software in detecting academic plagiarism was evaluated by copying various texts on the subject of science from randomly selected online platforms. The obtained results are presented in the following table:

Figure (5): Table showing the results of similarity measurement for different texts using Turnitin.

Topic	Platform Copied From	Plagiarism / Citation Rate	Was the Source Detected?	Source Detection Rate	Match with Real Percentage (100%)
-------	----------------------	----------------------------	--------------------------	-----------------------	-----------------------------------

(Blank)	Mawdoo3	71%	No	28%	Somewhat matching
Science	Wikipedia	96%	Yes	00%	Highly matching
(Blank)	Islam Online	56%	Yes	16%	Not matching
Article	ASJP	13%	No	00%	Completely not matching

Analysis: The detection of academic plagiarism depends primarily on the content available in the software's database, the size of that database, and the frequency with which it is updated. The relationship between the software's efficiency and these variables is directly proportional: the larger and more comprehensive the database, and the more regularly it is updated, the higher the software's detection performance. Nevertheless, the decisive factor in identifying plagiarism remains whether the original source is included in the software's database.

The similarity index is not always directly related to identifying the actual source of plagiarism, particularly in the case of electronic resources that provide open access. A text may be copied from one online source, while the software identifies a different website as the source of the matching text. This occurs because the source from which the plagiarism was committed may itself have reproduced the text from another online source. Regardless of this, the essential objective is the software's ability to detect the existence of plagiarism, whereas accurately identifying the original source depends on several factors, including the originality of the text in the electronic source.

In each of the four experiments, the text was copied verbatim from its original online source. Despite this, the reported similarity percentages varied. In the first case (Mawdoo' platform), the reported similarity closely approximated the actual percentage of copied content. In the second case (Wikipedia), it was almost identical to the true similarity. In the third case (IslamOnline), the reported percentage differed noticeably from the actual value. In the fourth case (ASJP), however, the reported similarity was entirely inconsistent with the actual percentage. The researcher explains these findings as follows:

Regarding the ASJP platform: The software detected only the abstract, keywords, and title of the article as matching content from ASJP, whereas it failed to identify the body of the article as having been copied from the same platform. Consequently, the similarity percentage shown in the table corresponds to quotations cited from other sources rather than to the copied article itself. This may be attributed to the software's inability to access the full text of the article, despite successfully accessing the abstract, title, and keywords hosted on the same platform.

Regarding the other platforms: The observed discrepancies may be due to the possibility that these platforms had themselves paraphrased texts obtained from other electronic sources. Another possible explanation is the continuous updating of the content available on these platforms.

5.2 Testing the Software's Efficiency in Handling Different File Formats

The software's efficiency in processing different file formats was evaluated by submitting the same content in different formats. The results are presented in the following table:

Figure (6): Results of similarity measurement for the same content submitted in different file formats using Turnitin.

Input Content	Content Format	Plagiarism / Citation Rate	Plagiarism Source Detection	Match with Real Citation Percentage
Article from	Written text	1%	No	Completely not matching
Platform	Word	1%	No	Completely not matching
ASJP	PDF	0%	No	Completely not matching

Analysis: The results presented in the table demonstrate the software's ability to process data from different input formats, reflecting its high efficiency in handling the formats examined. The similarity percentage remained identical across all formats except PDF, for which a lower similarity percentage was recorded. This difference can be attributed to the PDF reader integrated into the software, which failed to recognize certain expressions contained in the file. Consequently, those expressions were not identified as matching text.

For this reason, it is generally recommended to upload files in Word (.doc or .docx) format when submitting documents for similarity analysis.

The table also illustrates the software's failure to identify ASJP as the original source of the copied article. The reported similarity percentage is attributable to quotations cited from other references or to the article's metadata namely the title, abstract, and keywords which appear on the platform's first page when the article is searched. It does not indicate that the software recognized the article itself as having been copied from the platform without modification.

5.3 Testing the Software's Efficiency in Detecting AI-Generated Texts

Figure (8): Table illustrating the efficiency of Turnitin in detecting plagiarism in AI-generated texts.

Text	Topic	Word Count	Citation / Plagiarism Rate	Source Detection (ChatGPT)
Text 01	E-Learning	367	09%	Not detected
Text 02	Intellectual Property	328	04%	Not detected
Text 03	School Dropout	360	00%	Not detected

The software's efficiency in detecting plagiarism in AI-generated texts was evaluated by measuring the similarity index of three texts generated using ChatGPT based on artificial intelligence techniques. As shown in the table, the software failed to identify ChatGPT as the source in all three cases. The reported similarity percentages were attributable only to incidental similarities between the generated texts and materials available from other sources.

Although the generated texts dealt with well-established and extensively researched topics, the artificial intelligence system produced texts that either differed substantially from existing sources or exhibited only limited similarity to other texts addressing the same subjects.

The researcher attributes these findings to the fact that the software does not rely on artificial intelligence for plagiarism detection or similarity analysis. Instead, it depends primarily on its database by comparing textual sequences and calculating the frequency of matching expressions.

5.4 Testing the Impact of the Characteristics of the Arabic Language on the

Figure (9): Table illustrating the impact of the characteristics of the Arabic language on similarity measurement.

Text Status	Number of Paragraphs	Word Count	Citation / Plagiarism Rate	Nature of Edits	Count
Text Before Editing	<i>(Blank)</i>	<i>(Blank)</i>	<i>(Blank)</i>	Reordering / Rearranging	24
During Editing	03	293	91%	Deletion	17
Text After Editing	03	260	00%	Direct Mention	10
<i>(Blank)</i>	<i>(Blank)</i>	<i>(Blank)</i>	<i>(Blank)</i>	Synonymy / Paraphrasing	09
<i>(Blank)</i>	<i>(Blank)</i>	<i>(Blank)</i>	<i>(Blank)</i>	Derivation	15

The impact of the characteristics of the Arabic language on plagiarism detection was examined by measuring the similarity index of a text copied verbatim from a Wikipedia article on science, both before and after introducing minor modifications.

The table shows that, before modification, the similarity index reached 91%, a very high percentage that closely approximates the actual similarity of 100%, since the text had been copied verbatim without alteration. By contrast, after only slight modifications, the reported similarity index dropped to 0%, despite the fact that the proportion of modifications introduced into the text did not exceed 26%.

These findings reflect the remarkable flexibility of the Arabic language and its exceptional ability to express the same idea through numerous potentially unlimited linguistic formulations. This is especially evident considering that the experiment involved a topic that has been extensively studied. Moreover, the modifications were limited to changes in the structural organization of the original text rather than involving comprehensive paraphrasing, stylistic transformation, changes in attribution, metonymy, metaphor, or other rhetorical devices.

6. Conclusion

Intellectual property is the right to enjoy and dispose of intellectual production. Any infringement of this right constitutes an offense punishable under intellectual property protection laws, as well as copyright and related rights legislation. The advancement of scientific research depends

fundamentally on safeguarding this right. Based on the findings of the present study, the following conclusions may be drawn:

The importance of similarity detection and plagiarism detection software lies in its ability to protect intellectual property by handling massive volumes of data and the diverse means through which information is disseminated.

Similarity detection software identifies plagiarism by measuring similarity in textual expression (wording) rather than similarity of ideas, since ideas themselves cannot be measured. Consequently, these programs assess the manifestation of thought through human linguistic expression as a means of calculating the similarity index.

Considering similarity detection software as a complete means of protecting intellectual property remains relative, because a single idea can be expressed through numerous different formulations. In such cases, intellectual rights may be infringed without the software being capable of detecting or protecting them.

Relying exclusively on this approach means that effective protection of intellectual property can only be achieved after all possible linguistic expressions of the same idea have been exhausted a delayed form of protection that may result in substantial losses and the erosion of intellectual rights.

Similarity detection software operates by measuring the number of instances in which sequences of words are repeated or matched.

The efficiency of similarity detection software depends on the size and quality of its database. Since these factors vary from one program to another, similarity reports may differ accordingly.

Because no database can encompass all available information, plagiarism detection inevitably remains relative. Its effectiveness depends primarily on whether the original source exists within the software's database, as well as on the software's ability to process different document formats.

Including references and bibliographic sources in the similarity calculation is unfair to researchers, as they are penalized for material that cannot legitimately be altered.

Establishing an upper limit for the acceptable similarity index in scholarly research encourages innovation and originality.

Owing to its distinctive linguistic characteristics, the Arabic language is remarkably flexible, allowing a single idea to be expressed through an almost unlimited number of formulations. This flexibility presents a significant challenge to the ability of similarity detection software to identify plagiarism and protect intellectual property.

Linguistic features such as fronting and postponement (word-order variation), omission and addition, synonymy, and derivation reduce the effectiveness of similarity detection methods based primarily on matching consecutive sequences of words.

Arabic is a root-and-pattern (derivational) language, whose morphology relies on roots and the affixes attached to them. This characteristic further complicates plagiarism detection, particularly with respect to grammatical variations such as verbal inflection, singular, dual, and plural forms, masculine and feminine forms, and active versus passive constructions, among many others.

To protect intellectual property and encourage innovation, the software imposes an important constraint on researchers when rewriting texts that exceed the permitted similarity threshold. It is not sufficient merely to produce wording different from the original source; researchers must also avoid formulations that may resemble those found in any other existing sources.

Recommendations

The researcher recommends the following:

Expanding the software's database by incorporating as many academic journals, books, electronic platforms, websites, and other knowledge repositories as possible in order to achieve more comprehensive and accurate similarity assessment.

Integrating artificial intelligence into similarity detection software to improve its effectiveness in identifying plagiarism, particularly in texts generated using AI-based systems.