Globethics Repository



Using plagiarism detection software : the other side of the coin

This page was generated automatically upon download from the Globethics Repository. More information on Globethics see https://www.globethics.net. Data and content policy of Globethics Repository see https://repository.globethics.net/pages/policy.

Item Type	Book chapter
Authors	Eck, Nadine
DOI	10.58863/20.500.12424/4273122
Publisher	Globethics Publications
Rights	Globethics Publications;Attribution-NonCommercial-NoDerivatives 4.0 International
Download date	2026-11-04 23:23:54
Item License	http://creativecommons.org/licenses/by-nc-nd/4.0/
Link to Item	http://hdl.handle.net/20.500.12424/4273122

18.

USING PLAGIARISM DETECTION SOFTWARE: THE OTHER SIDE OF THE COIN

Nadine Eck

Abstract

The conclusions of this article are the result of a study conducted over three years, based on the expertise files that the author established as a scientific collaborator of the current IRAFPA. The use of similarity detection software was systematic for each case. The aim of this article is to demonstrate the absurdity of a persistent belief in universities: that it would be sufficient to call on the services of a computer services company specialising in so-called "anti-plagiarism" software to curb such cases. We will show, by example, what can and cannot be expected of them, and then we will compare the two most widespread in France, Urkund and Compilatio.*

* Corresponding authors: N. Eck. To quote this chapter: Eck, N., "Using Plagiarism Detection Software: The Other Side of the Coin" in: Bergadaà, M., Peixoto, P. (Eds.), *Academic Integrity: A Call to Research and Action*, Geneva: Globethics Publications, 2023, pp.435-455, DOI: 10.58863/20.500.12424/4273122 © Globethics Publications. CC BY-NC-ND 4.0. Visit: https://www.globethics.net/publications

1. Introduction

Similarity detection software is essential to prevent the massive fraud that occurs in uncontrolled situations. These packages do not claim to make the phenomenon disappear by detecting all cases of plagiarism, even though their functionalities are progressing at the same pace as the plagiarists' tricks. For example, we have often heard people accused of negligence for not checking work that is the subject of a dispute say, 'The plagiarism detection software did not see any fraud, so the thesis or article is not plagiarized.'

This is often an illusion, as it is very easy to use simple free paraphrasing software to reword an entire text. Some sites offer text 'rewriting features' that can make plagiarism undetectable by software.[465] Here again, only a specialist's analysis will be able to identify whether or not there is plagiarism by examining suspicious repetitions.

As for PhD theses, which are too long to paraphrase by this process, there are multiple specialized sites that clearly announce their intention: 'The paraphrasing service is a very special and sensitive service; for this reason, our linguistics team assists and provides the best paraphrasing service online. Plagiarism is bad and illegal. Worse, it can get you expelled! On our paraphrasing platform, you can benefit from the (best) paraphrasing prices...'.[466]

Let us be frank: when authors of works work undercover and are determined to deceive the reader, only a rigorous expert analysis can reveal plagiarism. This analysis must be carried out by a specialist in the discipline concerned: a reviewer for a scientific journal, an instructor, or a thesis director.

---

[465] For example, Smodin, *Réécrire, reformuler ou reformuler du texte | changeur de mots gratuit | français,* 2020; Le Reformulator, *Reformuler, réécrire ou reformuler du texte | Changeur de vocabulaire...,* 2019.

[466] Protranslate, *Service de paraphrase,* 2020.

For more than ten years, IRAFPA's analysis protocols have been guided by one mission: to enable the creation of a file that meets the requirements in terms of reliability, validity, and rigor that are indispensable to both investigative commissions and legal services that may be involved and, of course, to victims and whistleblowers. In the following pages, we describe the use we make—or do not make—of similarity detection software in the factual evidence files we compile at the request of claimants, victims, legal services, chairmen of commissions of enquiry, and whistleblowers.

2. The operating profile of knowledge delinquents

Plagiarists are like art forgers: they all have a specific modus operandi and, when we discover their style on a piece of work, they rarely change it. So our first job is to discover that modus operandi.

Thus, when a claimant sends us a list of texts declared to be plagiarized, a first analysis of the file consists in studying the admissibility of the claim on the basis of the evidence submitted to us. In the second step, the designated texts are subjected to software analysis, which in most cases allows us to detect other similarities. Exhaustiveness is not a necessary condition for a rigorous demonstration. Indeed, because the samples are representative of the whole text from which they are extracted, the results can be generalized to the whole text analyzed. A random search of samples is sufficient, especially when it concerns a work or an article of a certain length.

In order to establish the profile of the alleged plagiarist, we have defined the following five operating methods, together with Michelle Bergadaà.[467] They allow us to unambiguously describe the nature of the

[467] M. Bergadaà, *Le Plagiat académique. Comprendre pour agir* (Paris: Éditions L'Harmattan, 2015), pp. 57-59.

academic plagiarism observed and to make our reports as objective as possible.

Operating method 1: Textual repetition without elaborate masking. The main techniques used are verbatim plagiarism or copy-pasting of one or more sentences or expressions without quotation marks and without quoting the source; translation of texts written in a foreign language without quotation marks and without quoting the source; and summary of a text without quoting the author. In all three cases, however, the source author may be cited, but incorrectly referenced and/or cited in an inappropriate place, such as before the plagiarized passage or just at the beginning of it, so that readers think they are reading an original development when in fact they are reading plagiarized words.

Operating method 2: Masking process using simple or relatively complex techniques, which may be combined. The main ones are: reduction of the plagiarized text, partial summary, synonymy, and alternation of verbatim plagiarism and paraphrases; moving or interchanging words, expressions, propositions, or sentences (in the Molière style: *'Belle marquise, vos beaux yeux me font mourir d'amour'/'Vos beaux yeux, belle marquise, d'amour me font mourir'*[468]); moving a footnote into the body of the text (or vice versa), occasionally adding personal notes, small changes in the model (1850 becomes 1849), changing the tense, mood, or voice of verbs, changing the mode (affirmative, negative, interrogative) of a sentence; and referring to sources other than the plagiarized source—these other sources may themselves be plagiarized—so as not to arouse suspicion of plagiarism. These techniques are used when the source author is not cited or is cited inappropriately.

---

[468] Beautiful marchioness, your lovely eyes make me die of love/Your lovely eyes, beautiful marchioness, of love make me die.

Operating method 3: Disguise using sophisticated techniques. In addition to the techniques listed under methods 1 and 2, more sophisticated techniques may be used to make plagiarism particularly difficult to detect. This involves making a mosaic using various compositional methods. One of the most common is to create a patchwork text (the length of a page, for example) made up of snippets of text borrowed either from several source authors or from various passages of the same source author located in different places. Conversely, a continuous text from one source author may be broken up into snippets scattered in various places in the plagiarized text, where it may even be combined with snippets from another source author. This artificial collage, made up of elements that are necessarily, to varying degrees, decontextualized and detached from their original logical arrangement, produces, at best, a clumsy, obscure, or partially incoherent text, and at worst, incomprehensible gobbledygook that can nevertheless fool readers with the glitter of enigmatic newspeak.

Operating method 4: Appropriation of the thought/expression of a renowned author. This is a kind of plagiarism that can take the form of a whole book. The plagiarist copies or paraphrases a renowned author whose thought processes and stylistic characteristics they have fully assimilated. This identical reproduction, on a scale that can be large, is done without indicating the source author and/or by displaying adherence either to the model copied or to the school of thought represented by this model (the plagiarist will say, for example, that he is 'Bourdieusian' or 'Popperian').

Operating method 5: Appropriation of non-literary data. This modus operandi, which can be practiced in all fields of knowledge, particularly affects the so-called 'hard' sciences. It consists of appropriating data of all kinds that do not fall within the scope of literary expression understood in the broad sense. These data include, among others, demonstrations, developments, passages formulated in specific

languages (mathematics, physics, etc.), diagrams, curves, graphs, tables, plans, illustrations, drawings, maps, photographs, etc. Masking procedures comparable to those used in methods 1 to 3 may be used. The source author is not cited.

3. Detection software: useful or a pipe dream?

It is easy to see that, of the operating methods described above, only type 1 plagiarism will definitely be detected; the software may potentially alert the user to a procedure under method 2. Similarity detection software is essential to prevent the massive fraud that occurs in uncontrolled situations, but it would be wrong to believe that such software is intended to eradicate the problem and steer knowledge delinquents who operate in our academic world back onto the 'straight and narrow', be they students or renowned researchers.

To be convinced of this, you just have to ask for a free trial, if your institution does not subscribe to one of the plagiarism detection software application. The procedure is generally as follows: you identify yourself by indicating your email address. You then receive your access codes, in order to use the document analysis service. You select one (or more) document(s) to be analyzed and/or compared. The document is loaded into the workspace available to you. The analysis of the document is launched according to the documents entered, as well as documents referenced on the Web, those saved by users of the software (with their agreement), and even publications for which some software applications have subscriptions.

The software's response time varies greatly. In addition to the fact that some software works very fast (a few minutes) and others much slower (a few hours or even days for the slowest), the 'weight' of the documents to be analyzed and the method used are determining factors. Software to be installed on a personal computer should in theory work faster but it requires the computer to be switched on throughout the

analysis. On the other hand, online software analyzes the documents while the computer is switched off: everything happens remotely on the software's server.

We receive exhaustive and precise results such as the percentage of similarities of the analyzed document, the set of passages similar to other documents (highlighted in color), the exact references of the sources with similarities, and, finally, a list of similar sources sorted by probability.

The analysis work really begins with the results displayed by the software. For, on the one hand, similarity is not synonymous with plagiarism of protected works (legitimate quotations and some very common expressions, for example, are identified as similarities by the software, but obviously do not reveal an act of plagiarism); on the other hand, the presence of seemingly minor similarities does not mean 'absence of plagiarism'. And this is where the greatest confusion occurs. Many people do not need to do the checking work because the software did not detect anything. The percentage of similarities between the analyzed document and the sources found is, in fact, only an indicator whose importance varies according to the modus operandi used by the plagiarist.

The percentage of similarity is calculated according to the amount of *authentic* text compared to the amount of *similar* text found. Most software indicates which passages are similar so that one can judge for oneself whether they are quotations or plagiarism. Secondly, the software indicates which sources are found for each similar passage. Some tools even classify the most frequently found sources.

Some software builds its own database by indexing websites and documents found on the Internet or submitted by users. Other software programs are meta search engines: they do not build their own document database but query and centralize the results of the search tools available on the Internet. Finally, some software mixes these two technologies. In

all cases, content that is freely accessible on the Internet is detected. In order to have access to publications for which access is fee-based, anti-plagiarism software publishers must establish specific partnerships with the distributors of these documents.

When an analyst reads the result provided by the similarity detection software and finds that an extract is not plagiarized, but quoted, and also finds that many extracts come from sources that she herself has recommended to her students, what should she do? First of all, she needs to have access to an interface that allows her to modify the initial result. Depending on the technology used, some software allows the analysis results to be adapted and adjusted. It is possible, for example, to ignore a given source or extract and recalculate the percentage of similarities.

The analysis can be complemented, to a lesser extent, by the use of search engines, which can identify fraudsters, at least those who have taken advantage of their teacher's naiveté by copying easily accessible data. The truth is that search engines have significant shortcomings: they do not have access to the contents of sites offering to sell rewritten work in its entirety. Nor do they have access to password-protected pages. Finally, it is important to know that Wikipedia pages, for example, are regularly modified and that the software is not able to find deleted content. Furthermore, it is often enough, for example, to replace a single space with a double space between two words or, of course, to simply reverse the words for the search engine not to recognize a quotation. Thus, most software detects only plagiarized excerpts and not translated passages. Machine translation software exists, such as Deepl, which is based on artificial intelligence and is incredibly efficient.

Thus, if the work of educating people about citation ethics has not been carried out by the team of supervisors, colleagues, or editors in the publishing world, it is necessary to carry out an *a posteriori* assessment, based on the documents submitted or published. The work of a specialist in the discipline is then essential to detect anything that has fallen

through the cracks of the software. And the cracks can be quite wide, as we show below.

4. Assessment of software use in expert reports

IRAFPA regularly receives cases requesting expertise and mediation. In 2016, we compiled an exhaustive table of the cases investigated in order to determine the real value of similarity detection software. We used two programs, whose functionalities are described in section 5. Table 1 summarizes the fifteen cases we examined over a given period. For half of the cases analyzed, the similarity detection software was insufficient or inoperative. For the other half, the picture is mixed. Thus, whenever academic managers claim that their institutions have put in place integrity enhancement measures and mention the use of software as clear evidence, we can make this table available to them.

Table 1. Software evaluations of fifteen cases of potential plagiarism.

Case and discipline	**Relative usefulness**	**Comments**
Case 1: Medicine	Allowed plagiarism to be detected and proven	The software detected a significant percentage of the plagiarism as it was mainly verbatim plagiarism, from English to English.
Case 2: Law	Allowed plagiarism to be proven	The software detected a significant percentage of plagiarism since the plagiarism process consisted of frequent verbatim plagiarism in the same language, French.
Case 3: History	Allowed plagiarism to be detected	The software detected a significant percentage of plagiarism, as there was mainly verbatim plagiarism, from English

		into English. But the large number of sources and the long period covered by all the disputed texts made the examination very cumbersome, even with the support of the software.
Case 4: Computer Science	Allowed plagiarism to be detected	The software detected a significant percentage of plagiarism as there was essentially verbatim plagiarism. However, it did not detect plagiarized tables and graphs.
Case 5: History	Allowed plagiarism to be detected	The software detected a percentage of plagiarism from another source, which was cited, but incorrectly. The 'mechanical' analysis of an expert in the field revealed other frauds (invented data, falsified fieldwork, etc.).
Case 6: Law	Allowed plagiarism to be detected	The software detected sources of plagiarism, but the analysis had to be conducted 'mechanically' by an expert in the discipline, as paraphrasing was dominant and several source authors were used.
Case 7: Management	Use of two software programs that detected several plagiarized passages	The result given by the software had to be refined: a specialist analysis was needed to flesh out the corpus of plagiarized texts. Some of the plagiarized passages were cut by a few sentences or did not appear in the same order in the thesis and in the original article. Nevertheless, it was mainly verbatim plagiarism.
Case 8: Economics	The analysis by two software programs did not give convincing	The plagiarism involved quoting sources in an ambiguous way to create confusion between what was original and what came from

	results: very low rate of similarities	the source authors' research. This modus operandi, associated with the use of occasional masking (use of synonyms, paraphrases, change of verb tenses), is difficult for software to detect.
Case 9: Geology	The software produced inconclusive results	The plagiarism involved the translation of documents, and disguised figures and tables that are not detectable by similarity detection software.
Case 10: Sociology	The software produced inconclusive results	The plagiarism involved repetition of other people's ideas and paraphrases, but mainly self-plagiarism.
Case 11: Anthropology	Software did not work	Translation and/or paraphrasing of a French text into English: a process that cannot be detected by software.
Case 12: Literature	Software did not work	The plagiarized books and articles were not accessible on the Internet or were accessible through a paid service. The software does not give any results in this case.
Case 13: Finance	Software did not work	The plagiarism involved translation from French to English of part of the thesis, copy and paste for some tables, and manipulation and change of results for others. The software is not efficient when plagiarism involves translation or copying

		tables.
Case 14: Anthropology	Software did not work	The plagiarism involved translation from French to English, verbatim plagiarism, and paraphrasing. Apart from the verbatim plagiarism, these processes cannot be detected by similarity detection software.
Case 15: Geography	Software did not work	The plagiarism involved manipulation of maps and tables and modification of photos. The software does not take these elements into account.

Despite the obvious limitations, this kind of software is now a must. The question now is how can we best support researchers and teachers? In the following section, we will compare two programs that we have used.

5. For those who want more information: comparison of two programs

We would like to thank the two software publishers—Compilatio and Urkund—which kindly agreed to take part in the comparison and answer our questions.[469]

Compilatio is a French software.

In 2005, teachers in France expressed their needs for plagiarism control to the managers of Six Degrés, a company

[469] The full table is available on the IRAFPA website: https://irafpa.org/en/methods/use-of-software-programs/a-comparison-test/.

> specializing in web design. The developers and the teaching staff brainstormed together on the solutions to be considered. Frédéric Agnès, one of the two partners at Six Degrés, decided to take on the project. The first version of Compilatio was released in 2008. In 2009, the team working on Compilatio created a new company of the same name, integrated into the Six Degrés holding company.[470]

Compilatio performs a three-level comparison: open-access Internet, documents deposited at your university, and documents deposited by all Compilatio users (respecting the confidentiality of documents). Depending on the wishes of the client organization, it is possible to add archives of student work from previous years or collections of documents that you can transmit via the software. Users can also add any document available to them to their own 'reference library' at any time. Compilatio can analyze all documents written in the Latin alphabet, in all languages, but it cannot compare texts from two different languages. However, Compilatio has been working on an algorithm for this purpose for several years.

With Compilatio's license, there is no restriction on the number of documents whose content is in the 'reference library'. However, there is a storage quota for the original files of the documents analyzed by the users, depending on the level of service chosen. It is possible to analyze as many documents as desired, without restriction in the context of individual use, in a normal academic context. Finally, Compilatio ensures confidentiality by offering the possibility to completely delete the documents of one's choice, without allowing sharing or external access. For example, if one chooses to keep the content of student work confidential, Compilatio will not provide copies of documents to third parties.

---

[470] A. Hamel, 'Comment utiliser un logiciel anti-plagiat?', *Thot Cursus,* 24 October 2011.

Urkund is a Swedish software.

> Urkund was born in the academic world. A team of researchers had the idea of a networked service that could help them detect and deter plagiarism, hence the birth of Urkund in the autumn of 2000... Urkund has continued to grow and develop over the years and has become Sweden's leading anti-plagiarism service... Urkund is owned and developed by PrioInfo AB. PrioInfo is a company that has been meeting the demands and needs of information-intensive companies for over 25 years.[471] Urkund compares your documents with all the sources available on the Internet, 45 billion websites, the documents Urkund has already received, in the archives, about 17 million documents (as of 15 February 2016), the publications accessible on our partners' databases, i.e. 4,000 news sources, a database of more than 1,000,000 newspapers.[472]

Urkund can analyze documents in all languages that use the Latin alphabet 'and has the possibility to analyze Arabic, Mandarin, Hebrew'. However, it cannot compare texts from two different languages.

With the Urkund license, it is possible to analyze any number of documents and to support any number of users (teachers). An unlimited number of documents can be saved without size limits. Urkund ensures the confidentiality of certain documents by completely deleting them without allowing sharing or external access. If you choose to keep the content of student work confidential, Urkund will not provide any copies of documents to third parties. At the end of the contract, Urkund can return all data to the university and then destroy all stored files.

---

[471] In September 2020, Urkund became Ouriginal (a synthesis of Urkund and PlagScan). For more information on the merger between Urkund and PlagScan, see Ouriginal, *Our Story,* 2020; see also Urkund, *Le système Urkund,* 2020.

[472] Urkund, *Le système Urkund.*

5.1 Ergonomics

Compilatio is intuitive and easy to use, but offers fewer features than Urkund (e.g. simultaneous access to other sources where there are similarities). Urkund allows simultaneous access to many features on the same page, but its presentation is more complex and sophisticated.

5.2 Displaying similarities

With Compilatio, the text of the analyzed document and of the source appear in their entirety: similarities not recognized by the software and not marked in color can be detected by instructors during their analysis. The detection of paraphrases or sophisticated plagiarism is also facilitated by the fact that the two texts appear opposite each other. Words in bold (red color) mean that the similarities are exact. In contrast, with Compilatio not all verbatim sentence fragments are colored; the analysis requires more time. Also, some words shown in color are not verbatim plagiarism.

With Urkund, the display allows simultaneous access to other sources where the same similarities in a sentence have been detected. 'Urkund always shows the best source on the page, the one closest to the text, but also refers to 5 other sources. These other sources are considered alternative and are indicated in the left margin.'[473] In contrast, the text of the analyzed document appears in its entirety, but only the similarities detected in the source text appear opposite it. It is therefore not possible to know what the software did not recognize or what is a paraphrase: it is necessary to click on the source link to consult the original article. The analysis is therefore more time-consuming, especially as the two texts are no longer facing each other. The analysis is also less easy, as Urkund converts all characters to Verdana (this is the case, for example, for the detection of chapter headings): 'During the analysis process, all italics, highlighting, and bold have their fonts

[473] Translation from the URKUND Administrator Guide.

replaced by Verdana to facilitate single screen review. All tables and images that cannot be converted to text have also been removed.'[474]

5.3 Similarity percentages

With Compilatio, the document submitted for analysis is broken down into a number of 'parts', depending on the length of the text submitted. Compilatio presents an overall percentage of similarities for the whole text, but also a percentage by 'parts'. The percentages are given in relation to the document analyzed: 28% similarity, for example, means that 28% of the text contained in the document submitted for analysis was recognized as similar to the sources. For each source, a specific percentage is announced, which means that the analyzed document contains X% of text similar to that source. The set of similarities of an analyzed document is composed of the summed similarities for each source. It is possible to designate sources that you do not wish to take into account so that they are ignored when calculating the similarity percentage (to do this, simply check the selection box next to the desired source, then click on the 'ignore' button). All other sources, whether 'very likely' or 'unlikely', will be taken into account in the calculation of the similarity percentage (text areas that have the same similarities to several sources are only taken into account once). The result is a similarity percentage for each part and a similarity percentage for the whole document.

The advantage is that the selections made to refine the analysis (removal of sources or not) do not disappear after the software is closed. They can nevertheless be modified at any time by a simple click. Passages in quotation marks can easily be ignored for the calculation of percentages. In this case, simply answer the question: 'Exclude text in quotation marks from the percentage of similarity.'

[474] Translation from the URKUND Administrator Guide.

Note that with Compilatio, percentages, except in cases of verbatim plagiarism, are not indicative of the extent of plagiarism. They are only hints for further analysis—which is necessary—since paraphrasing, sophisticated verbatim copying, verbatim translation, and graphics, images, and non-textual data, on the whole, are not detectable by the software.

Urkund provides: (1) an overall percentage of similarity: 12% similarity, for example, in a 700-page text means that 12% of the document submitted for analysis is identical to all the sources found by Urkund; (2) a percentage for each source for which the software has detected similarities; and (3) within the same source, a percentage linked to each text extract where similarities have been detected. In the latter case, the percentage represents the degree of similarity, in detail, that the text shares with the source extract. This number helps to clarify the review process: 100% means that the text is identical to the source extract and 50% means that half of the words in the text differ in some way from the source extract. Similarities below 30% are not, in principle, highlighted. If a particular kind of information is considered irrelevant, it can easily be deactivated.

The overall percentage can be refined by ignoring pieces of text detected as similarities within the same source. However, although the result of the selections made (removal or otherwise of certain passages) can be sent by email, these selections disappear when the software is closed. If you save the Internet link, you can return to the selections made. Moreover, passages in quotation marks appear in color if you wish, but, in a passage including both verbatim copying and quoted passages, it is not possible to remove the parts in quotation marks so that they are no longer taken into account in the calculation of percentages.

It should also be noted that, with Urkund, percentages, except in cases of verbatim plagiarism, are not indicative of the extent of plagiarism. They are only hints for further analysis—which is

necessary—since paraphrasing, sophisticated verbatim copying, verbatim translation, and graphics, images, and non-textual data, on the whole, are not detectable by this software, just as they are not found by Compilatio. However, Urkund does show the words that differ between the two texts where the software has detected similarities (see section 5.6).

5.4 Detection of attempted manipulation by a fraudster

Compilatio can detect attempted manipulations, which are indicated by a pictogram. 'Compilatio Support' told us that new developments have been made to prevent the software from being bypassed, in particular the detection of unanalyzable text (triangle pictogram with an exclamation mark, which means that part of the document may potentially have been modified to avoid source detection). Urkund can also detect attempts at manipulation, which are indicated by 'Warnings'. Warnings also detect the manipulation of spaces (with the addition of a blank letter, for example): 'We are also testing a new function, so that we can show what is in parentheses in the texts analyzed.'

5.5 Analysis reports

Compilatio offers (1) a 'summary' tab of the report, which displays a general overview of your document, with the top sources (main sources found) and the corresponding similar passages. You can access the website directly by clicking on the source. (2) A 'Full text' tab: your document is presented in full with the similarities found. (3) A 'Sources' tab of the report displays all the sources that are similar to your document, sorted by percentage and by degree of relevance. Some sources are indicated as belonging to 'another user': these are sources submitted by authors who have opted for anonymity or 'external sources', that is, documents from a Compilatio user outside your university. To preserve the desired anonymity, the data are encrypted, but Compilatio does show similar parts of the texts. This display is a

great help in the case of substantial plagiarism, especially since it is possible to obtain access to the document via Compilatio's services. The procedure is as follows: send Compilatio certain information (account name, file name, document name, source concerned), then wait until the person agrees to transmit the source in question and contact you, after Compilatio services have communicated your contact details.

The *Urkund* report presents the text of the document submitted for analysis with the similarities and source references transcribed in color at the point in the text where these similarities were detected. The corresponding percentages are also indicated. When two samples of similar texts appear side by side, the software provides the possibility of visualizing, in detail, the differences between the two texts. This is the case, for example, for words that are not found in one of the two texts, or differences in tense or synonyms, for example:

> when the *Show detailed text differences* button is *on*, the differences are indicated on the source side, in the form of colored rectangles (highlighting) on the words that differ from the document under review. This happens, for example, when a word is missing from the source but present in the document: here, the colored rectangle is empty or, if there is one or more sentences in the examined document that are not present in the source, or the presence of one or more words in the source which are absent from the examined document; it also occurs when there is a word in the source that also appears in the examined document, but in another form such as a synonym, a changed tense, a wrongly spelled word or a word that is similar to it. For example, *In some cases* becomes *In some circumstances.*[475]

[475] Translation from the URKUND Administrator Guide.

6. Conclusion

The study of the various cases of plagiarism that we have been entrusted with shows that similarity software is unable to identify sophisticated plagiarism or plagiarism relating to certain representations such as maps, tables, or photos. Furthermore, sources are not accessible in the following cases: when the author of the analysis chooses to remove a document from the 'Reference Library' (Compilatio); when the identified source is available online, but access to it is limited to those with access rights (e.g. password protection, because of subscription); of course, when the textual content of the source is not available online, such as books in hard copy; and when the source is no longer available at the time the analysis was carried out (e.g. Wikipedia articles, which are frequently updated). Under these conditions, software companies, aware of the absolute necessity to improve access to sources, negotiate partnerships with scientific journals and encyclopedias to expand their databases, with varying degrees of success.

Nevertheless, software is a tool that gives us access to indicators of the modus operandi used. On the basis of these indicators, and apart from the case of verbatim plagiarism, an analysis is essential because, on the one hand, the subterfuges used are increasingly sophisticated, and on the other hand, the software available on the Internet to make masking and translation easy is both accessible and easy to use. Thus, although similarity software is constantly being improved, at this stage of its development, it should be considered above all as an essential tool for raising awareness and deterring plagiarism.

Bibliography

Bergadaà, M., *Le Plagiat académique. Comprendre pour agir* (Paris: Éditions L'Harmattan, 2015).

Hamel, A., 'Comment utiliser un logiciel anti-plagiat?', *Thot Cursus,* 24 October 2011. https://cursus.edu/articles/20780/comment-utiliser-un-logiciel-anti-plagiat

Le Reformulator, *Reformuler, réécrire ou reformuler du texte | Changeur de vocabulaire...,* 2019. https://www.lereformulator.com/fr/reformuler-vos-textes-automatiquement

Ouriginal, *Our Story,* 2020. https://ouriginal.com/our-story/

Protranslate, *Service de paraphrase,* 2020. https://www.protranslate.net/fr/paraphraser/

Smodin, *Réécrire, reformuler ou reformuler du texte | changeur de mots gratuit | français,* 2020. https://smodin.me/fr/reformuler-automatiquement-le-texte-en-français-gratuitement

Urkund, *Le système Urkund,* 2020. http://www.urkund.com/fr/about-urkund/350-about-urkund