

# Citation-based Plagiarism Detection: New Citation Software Tool Hunts for Plagiarism

**Author**

Enago Academy

**Post Url**

<https://www.enago.com/academy/citation-based-plagiarism-detection-new-citation-software-tool-hunts-for-plagiarism/>



Plagiarism is a serious issue in higher education and in [academic publishing](#) and at times can be difficult to identify. The deliberate misrepresentation of one's written work by copying it from other authors is considered as serious misconduct. For this reason, much time, effort and money have gone into deterring and combatting plagiarism, as it threatens academic integrity. Today, sophisticated detection software programs can systematically analyze a text in detail for evidence of plagiarism. But what clues do they look for? Detection software companies like *Turnitin* or *SafeAssign*, take a submitted text and compare it to those found on the Internet and also in huge databases, looking for copied segments of words. Once finished, a report is obtained showing the percentage similarity of the text to known sources. The report also color-codes the suspected text for easier identification. The problem is that this approach, while effective at spotting the "cut-and-paste" jobs of students, doesn't do so well when the sentences and paragraphs are either translated or paraphrased and they ignore citations altogether.

Now there is a new tool, called “Citation-based Plagiarism Detection” (CbPD), which uses the patterns in citation usage to ferret out sophisticated forms of plagiarism. Bela Gipp developed this approach, which is similar to bibliographic coupling, and her doctoral thesis is an excellent source for full details and in-depth analyses of CbPD and its potential applications to root out academic dishonesty and help ensure academic integrity.

## How Does CbPD Work?

So, how does CbPD detect a plagiarized document, be it a college essay, thesis, or journal manuscript? Simply put, it aims to create for a document a language-independent citation-based digital fingerprint, in a process called “Citation Order Analysis” that involves three main steps:

1. CbPD takes a document and, according to a set of heuristics, identifies the actual pattern of citations within it, including their relative positioning.
2. These citations are then cross-checked with and linked to their presentation in the bibliography.
3. This is followed by the computation of a citation-based similarity among two or more documents.

Hence, the semantic fingerprint CbPD generates can be compared to many existing academic documents. Moreover, this can be done in two ways: *basic*, which considers only the order of citations, and *advanced*, which considers both the order and length between any two citations. To get around the problem of slight changes in the order of citations, CbPD relies on tolerance sequence analysis algorithms.

The key here is that CbPD is able to [check for citation](#) overlap vis-à-vis that of other documents, in part by nicely color-highlighting the matching citations and by also providing a central column graphic, displaying citation nodes and length, to visually examine the similarity between any two documents. CbPD testing is also fast, too, at 2 seconds per publication, and it has performed reliably well in a battery of test sets using hidden plagiarized documents.

## Don't Succumb to Rogeting

Apart from the many ghostwriting services available to those with the time, money, and absent ethics, aspiring plagiarists often employ other strategies to circumvent detection software. In addition to translations and [paraphrasing](#), one growing strategy is to use synonyms in the place of original words, made easy via the thesaurus that is available in many word-processing programs as well as online free services. By engaging in such “Rogeting”—the name given by Chris Sadler at Middlesex University, in reference to the decades-popular editions of ‘Roget’s Thesaurus’—using enough synonyms will mask the originally copied text to confuse the detection software. However, given the nuances of meanings and many words in English, this tactic can be disastrously embarrassing, resulting in strange wording (e.g., “wise saying” becomes “clever motto”) if not outright gibberish. Thankfully, an experienced teacher can spot Rogeting. All writers should

learn to use a dictionary, and avoid jargon whenever possible.

## General Concerns

Notwithstanding the degree by which one can plagiarize, the criticism of detection software programs goes beyond just the false negatives (up to 39% for *Turnitin*). Where does plagiarism start and end, when given a similarity type of score? Might CbPD also be prone to pesky false positives? Any detection software, including CbPD, has inherent limitations and should not be used in a draconian way, no matter how convenient. Many students cheat, surely, but others may be hapless or careless, or simply ill taught on how to properly cite material. Another concern is that after knowing how CbPD functions; will this result in craftier ways to disguise plagiarism to avoid getting caught? For example, as is now possible by altering the binary text code in Rogeting.

## Benefits of CbPD

In short, using CbPD in tandem with other detection software programs should make plagiarism easier to detect. This should discourage academic dishonesty and promote academic integrity. Detecting gross scientific misconduct, as in the case of the revoked doctoral thesis of Karl-Theodor Zu Guttenburg, is an auspicious victory. The major contribution of CbPD is its independence: its efficacy does not depend on how words are strung together or their language. This is because it analyzes the citation order in a document and overcomes limitations of word-based detection software, which is often stymied by translations, paraphrasing, and Rogeting while requiring comparatively minimal computational complexity to do this. Finally, CbPD has a broad appeal and is applicable to any kind of document that has citations in it.

## Future of CbPD

Unfortunately, the incentives to plagiarize remain strong. Detection software is widely used and popular in higher education and journal publishing, but CbPD is not a panacea. This novel tool should not substitute but rather supplement the word-based plagiarism detection software programs currently available. CbPD should also be used cautiously and judiciously to catch plagiarism. Nevertheless, with its early victories, CbPD may indeed soon prove itself a formidable hunter of plagiarism in academic research.

## References

1. Mathieu Bouville (2008, March 11) *Plagiarism: Words and ideas*. Retrieved from: <https://arxiv.org/pdf/0803.1526>
2. Carl Straumsheim (2015, July 15) *What Is Detected?* Retrieved from: <https://www.insidehighered.com/news/2015/07/14/turnitin-faces-new-questions-about-efficacy-plagiarism-detection->

[software?utm\\_source=Inside+Higher+Ed&utm\\_campaign=6d648899c0-Insider\\_Update\\_201508&utm\\_medium=email&utm\\_term=0\\_1fcbc04421-6d648899c0-198554481](https://www.enago.com/academy/software?utm_source=Inside+Higher+Ed&utm_campaign=6d648899c0-Insider_Update_201508&utm_medium=email&utm_term=0_1fcbc04421-6d648899c0-198554481)

3. Cory Turner (2014, August 25) *Turnitin And The Debate Over Anti-Plagiarism Software*. Retrieved from:  
<http://www.npr.org/sections/ed/2014/08/25/340112848/turnitin-and-the-high-tech-plagiarism-debate>
4. SciPlore. *Citation-based Plagiarism Detection*. Retrieved from:  
<http://www.scipline.org/projects/citation-based-plagiarism-detection/>
5. Bela Gipp and Jöran Beel (2010, June) *Citation Based Plagiarism Detection – A New Approach to Identify Plagiarized Work Language Independently*. Retrieved from: <http://www.scipline.org/wp-content/papercite-data/pdf/gipp10c.pdf>
6. Bela Gipp (2013, September 2) *Citation-based Plagiarism Detection*. Retrieved from: <http://www.scipline.org/wp-content/papercite-data/pdf/thesisbelagipp.pdf>
7. Rebecca Schuman (2014, August 14) *Cease Rogeting Proximately!* Retrieved from:



## Cite this article

Enago Academy, Citation-based Plagiarism Detection: New Citation Software Tool Hunts for Plagiarism. Enago Academy. 2017/04/21. <https://www.enago.com/academy/citation-based-plagiarism-detection-new-citation-software-tool-hunts-for-plagiarism/>