Plagiarism Detection Software

MISCONCEPTIONS

The stakes of ensuring originality of content have never been higher, both for authors and publishers. In the past year alone, public allegations of unchecked plagiarism have damaged the reputations of some of the most prominent scholarly publishers and scientific organizations. The National Science Foundation has launched an investigation of what may amount to nearly $100M in plagiarized grant proposals, and scientific journals report a tenfold increase in retractions over the past 20 years. Science writer Jonah Lehrer and renowned primatologist Jane Goodall both found themselves embroiled in plagiarism scandals related to forthcoming books, and German Chancellor Angela Merkel's Education Minister resigned after her PhD was revoked upon discovery of plagiarism in her doctoral dissertation.

Ready access to content online makes it easier than ever to reuse copy, whether intentionally or accidentally. However, increasingly sophisticated search technologies—notably Google—along with plagiarism detection software, are leading to faster and more complete discovery of acts of duplication.

Plagiarism detection software is accessible, affordable and effective, yet many common misperceptions exist about its use and limitations. This paper will explore seven of the prominent myths that surround plagiarism detection and provide guidance into how to best use these tools to ensure originality of an author's own work, or of work submitted for publication or grant application.

7 Misconceptions of Plagiarism Detection Software

Plagiarism Detection Software Automatically Detects Plagiarism

Plagiarism detection software, like search technology, is algorithm driven. With search technology, the algorithm is tuned to identify keywords; the intention is to surface keyword-matched content across crawled and indexed online content. With plagiarism detection software, the algorithm may be tuned to identify chains of linked words in a strict text-to-text match or it may be tuned to pick up more nuanced matches, as with paraphrasing.

The combination of technology and human analysis yields the most thorough plagiarism assessment.

But identifying matched content is only part of the process of detecting plagiarism. The remainder requires a critical human element—a thorough interpretation of the scan's results to determine whether the matched content constitutes an act of plagiarism or is simply a quote or excerpt. Search algorithms cannot parse intent, nor can they determine if matched content is properly cited. The combination of technology and human analysis yields the most thorough plagiarism assessment.

Plagiarism Detectors Are Inaccurate

Google tops the list of search engines for the breadth and speed with which it crawls and indexes content online. That said, content is growing at a pace that far exceeds Google's capabilities. In addition, content

that is gated or behind a pay wall will not be accessible to the Google search bots.

Good plagiarism detection software not only provides access to crawled and indexed online content, it can compare text to databases of gated journal and e-book content. The software may also compare submissions to a database of masters and doctoral theses as well as submitted paper content.

There is no way to be wholly comprehensive in indexing all content, but plagiarism detection software used in the proper context—for example in the review of journal submissions or student papers—is highly successful in identifying matched content by virtue of the inclusion of additional database content with which to search against.

Plagiarism Detectors Are Easy to Deceive

Especially as more academic institutions and publishers begin mandating the use of plagiarism detection software, there are increasing attempts to devise methods of tricking the system into passing off duplicated text as original. Students have gone to great lengths to cheat the system—substituting Cyrillic letters that look similar to Latin letters throughout the text, adding extra spaces between letters and words, or even adding "invisible" small letters in white font to differentiate the text, often posting their strategies online for others to use. What is less publicized is that behind the scenes, the creators of plagiarism detection software go to even greater lengths to anticipate cheating strategies and to incorporate fixes into their systems. Many plagiarism detection software systems automatically strip macros from submitted manuscripts and can easily detect oddly placed punctuation and invisible characters.

Software engineers estimate that in order for duplicated text to pass a plagiarism detection scan, the author would need to rewrite or revise every third word.

Most attempts to outsmart plagiarism detection software require effort beyond what is required to properly cite or paraphrase source material. Software engineers estimate that in order for duplicated text to pass a plagiarism detection scan, the author would need to rewrite or revise every third word.

All Plagiarism Detection Tools Are The Same

The key to a thorough plagiarism check is the database the software uses for comparison. Many popular services perform text searches of anything available on the Internet that isn't behind a pay wall. Others include a database of academic journals and archived papers. iThenticate compares submitted manuscripts to a database of 37 billion archived and current web pages, more than 92 million offline works from academic journals, magazines and research abstracts, and 37 million scholarly articles, books, and proceedings from tens of thousands of scientific and technical journals.

Additionally, many plagiarism detection services employ algorithms similar to those used in web search, providing only text-to-text matches. iThenticate utilizes a proprietary algorithm that transforms each submitted manuscript into a "digital fingerprint", which is compared—just like a human fingerprint—to an extensive database where subtle shades of similarity can be detected.

Plagiarism Detection Tools Are Only For Academia

While those in academia were the earliest adopters of plagiarism detection software, the utility of this technology has moved well beyond universities. Today, the range of industries employing plagiarism detection technologies is wide, and use of the software to ensure originality is standard protocol within many organizations.

Government agencies and granting bodies use plagiarism detection software to check grant proposals, legal documents, and financial reports for duplication and misconduct. Medical research institutions

screen manuscripts to avoid blatant plagiarism, duplication or self-plagiarism, grant misconduct and improper patient treatment. Scholarly journals and publishers rely on plagiarism detection tools to screen submitted work before publication to avoid copyright infringement and protect their reputations. Many publishers also suggest that submitting authors run their own scan prior to submission to catch accidental errors in citation or sourcing.

Plagiarism Detectors Are Only Useful for Uncovering Unethical Work

Plagiarism detection tools are designed to identify blatant instances of duplication but plagiarism comes in many forms, some of which are far more subtle and even accidental.

A good plagiarism detection software will locate not only exact matches in text, but can identify poor paraphrasing, missing citations and even grammatical errors. Researchers submitting a grant proposal or paper for publication may be juggling hundreds of citations, and it is easy to understand how a reference may be unintentionally omitted. It is these cases in which a plagiarism detection scan can make a meaningful difference, protecting a researcher's reputation and ensuring original, properly cited work.

Scholarly journals are also using plagiarism software to detect duplicative publication, sometimes referred to as self-plagiarism. Self-plagiarism may occur when a researcher submits an article to a journal before learning that it was accepted elsewhere or when an author borrows heavily from their own previously published work.

Plagiarism detection tools are commonly used to detect accidental duplication and common writing

While plagiarism detection tools can certainly detect the work of unethical writers, often they are used to detect accidental duplication and common writing mistakes.

Plagiarism Detection Software Is Time-Consuming

Editors at scholarly publications are facing record submission volumes, and researchers and academics experience tremendous pressure to publish. And while time is precious, most regular users of plagiarism detection software report that the additional time required for scanning—typically not more than a few minutes—is a worthwhile investment. A good plagiarism detection tool can even speed up the editing process by highlighting questionable passages so the author or editor can focus only on the most suspect parts of a work.

In a survey conducted by iThenticate, researchers reported the highest level of concern over plagiarism, with more than one in four reporting that plagiarism is a "very serious" problem in their field. The additional minutes spent ensuring that work is fully original prior to submission can allay a good deal of that concern. In that same survey, editors at scholarly publications reported routinely checking authors' submitted work for plagiarism—pre-checking gives authors added confidence through the submission process.

Conclusion

Of all of the ethical problems in academia and professional work, plagiarism is both one of the most serious and most common offenses. Plagiarism checkers can help spot plagiarism, both intentional and accidental.

Furthermore, with plagiarism detectors becoming more widely used, the likelihood of a document being checked at some point in the publishing process has increased significantly. Running a scan before submission or publication allows authors to remedy any errors without risking the reputation costs of plagiarism.