

# Plagiarism sleuths tackle full-text biomedical articles

October 25 2010

---

In scientific publishing, how much reuse of text is too much?

Researchers at the Virginia Bioinformatics Institute at Virginia Tech and collaborators have shown that a computer-based text-searching tool is capable of unearthing questionable publication practices from thousands of full-text papers in the biomedical literature.

The first step in the process is to find out what is restated before zeroing in on who may have crossed an ethically unacceptable threshold. The findings, published in [PLoS ONE](#), offer hope for curbing unethical scientific publication practice, a growing problem throughout the world.

"Building upon our earlier work reported in *Science* and *Nature*, which uncovered ethically questionable [journal articles](#) by comparing their abstracts, we have now re-tuned our computer program, eTBLAST, to scan thousands of full-text articles in PubMed Central, a freely available repository of full-text biomedical literature," said Harold "Skip" Garner, author on the paper and executive director of the Virginia Bioinformatics Institute. "Our goal was to measure how much and where in papers – for example, the introduction, methods or results sections – text is duplicated to establish the 'norm' in publishing. This will allow ethicists, which we are not, to begin to develop guidelines as to what is and what is not acceptable publication practice."

Although abstract search is an effective approach to detect potential plagiarism, full text analysis is needed to uncover all potential duplicate citations in the scientific literature. The researchers examined 72 011

full-text articles using the eTBLAST [computer program](#), which is only the tip of the iceberg for the number of published biomedical articles in the archives.

"We found that most papers are novel, as expected in scientific reporting, but even in papers reporting novel results, certain sections, such as the introduction or methods section, frequently have large amounts of content that appear elsewhere," said Garner. The researchers went on to explain that the re-use of text in certain sections, such as the methods section of papers, where authors provide details on how the work was done, is not a bad thing because it is important to use the accepted and most consistent techniques. "We also expect that other sections like the results section to be very unique just like the abstract. And this is the case in the overwhelming majority of papers," said Garner.

The current study revealed that the introduction section tended to be copied the most in similar citations. Also review articles were confirmed as being particularly prone to repetition.

"We believe this type of research will help us write better, more informative scientific papers, and prepare reviewers and journal editors for interpreting the similarity results that are emerging from the computational analysis of scientific papers. This approach is becoming increasingly commonplace as part of the scientific review process," added Garner. "Before crossing the line between acceptable and unacceptable writing, it is important to know the location of the line," concluded Garner.

**More information:** Link to the *PLoS ONE* article:  
[www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0012704](http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0012704)

Provided by Virginia Tech

Citation: Plagiarism sleuths tackle full-text biomedical articles (2010, October 25) retrieved 19 April 2024 from

<https://phys.org/news/2010-10-plagiarism-sleuths-tackle-full-text-biomedical.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--