

REVIEW

Rosalind Tedford
Wake Forest University

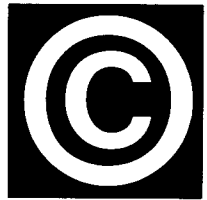
Plagiarism detection programs: A comparative evaluation

A 2002 survey by Rutgers' Management Education Center of 4,500 high school students found that "75 percent of them engage in serious cheating and more than half have plagiarized work they found on the Internet. [...] Some 50 percent of those responding to the survey said they don't think copying questions and answers from a test is even cheating" (Slobogin, 2002). These are the students that are coming in droves to our colleges and universities, and these are the students we are trying to impress in our writing classes and freshmen seminars with the definition and consequences of plagiarism.

As incidents of plagiarism rise on campuses across the United States, the question has been raised whether more students are actually plagiarizing, or if it is simply that more students are getting *caught* plagiarizing. The advent of the Internet certainly has made access to large bodies of pre-existing papers on every conceivable topic handier for our students. This access, in turn, gives students options ranging from saving a file, renaming it and turning it in as their own work, to assembling select sentences from multiple sources into one paper as easily as clicking a mouse.

The same Internet has also made detecting plagiarism much easier for faculty and administrators. Access and search methods that students use to locate papers are also available for use by faculty who grade their work. Often, typing a suspicious phrase or sentence into a Google search box can take a professor directly to the source material.

However, faculty have one additional tool at their disposal: plagiarism detection programs. These software programs and their services aggregate the papers available at the Internet "paper mills" and allow the user to search their content simultaneously. At many schools faculty and administrators are beginning to investigate plagiarism detection programs as both a tool for identifying plagiarism



and as a deterrent for students considering passing off others' work as their own.

Although some commercial plagiarism detection programs have been around for years, using technology for this purpose came into the spotlight in 2001. Lou Bloomfield, a Physics professor at the University of Virginia who became concerned over a student tip regarding suspected plagiarism, wrote a simple program to compare six-word strings in his database of papers from his Physics 105 and 106 classes. More than 1500 papers from three years of classes were compared—and 158 students were investigated on plagiarism charges (Artgetsinger, 2001).

In the end, only a small portion of these students was sanctioned, and the original authors of the plagiarized papers were not punished. "Of the 158 students whose cases were processed, 59 were formally accused of an honor offense. Twenty students, some of whom had graduated, were found guilty in trial, while the remaining 28 chose to forgo a hearing and withdraw from the University" (Quinlan, 2002). This Virginia story raised flags at educational institutions around the country regarding the level of plagiarism that was occurring. While the total number of students found to be plagiarizing was only 2% of the original number of students whose papers were compared, it still strained the University's judicial system and had repercussions across academia. Was this 2% an average or an abnormality? No one knew, because few if any studies had been conducted on such large numbers of students.

Into the void of apprehension created by lack of data that the UVA study spotlighted, came plagiarism detection programs and their services. Although some had existed for a while, it has been since 2001 that the number of schools using these programs appears to have gone up. People from academic institutions can barely gather together without someone mentioning that they have signed a contract with one of these program services or are investigating several to determine the best fit for their school. Institutions are spending a good deal of money to provide these applications to their faculty. But do they do what they claim and will



Rosalind Tedford is manager of Information Technology Center, Z. Smith Reynolds Library, Wake Forest University, Winston-Salem, NC 27109. Her e-mail address is tedforrl@wfu.edu.

they ultimately result in a decrease in plagiarism on campuses that use them? These are still unanswered questions in the longer run.

What plagiarism detection programs do

Most plagiarism detection program services have amassed large databases of papers from online paper mills. They compare papers submitted by faculty (copied and pasted into an online form, not submitted in their original format such as MS Word), and return reports that indicate passages where the submitted papers overlap with ones in their databases. Many of the program services also add submitted papers to their databases, although some do so only with faculty assent. In addition, many open access Web sites, such as the free portion of Britannica.com and other free Web encyclopedias and reference sites, are included in the databases.

Once a paper is submitted, the program returns a report to the instructor. Such reports usually rank a paper in terms of its similarity to other items in the database, and most also link the professor back to the source material so that she can see for herself if the similarities uncovered are alarming. It is then up to the professor to make the final decision about a particular paper and student author.

Some programs and services require software on the instructors' computers, others are purely Web-based. Some are free, some cost money. Some allow instructors to build up a database of their papers in the system so that comparisons can be made to past papers. Other features vary with the product.

What plagiarism detection programs do *not* do

These programs do not determine which portions of a paper have been correctly cited by the student. Consequently, some of the portions of a paper highlighted by the program may not be plagiarism if the student has cited their source. They do not compare papers to for-fee online databases of journal articles such as the ones subscribed to by libraries. If a student is quoting from an online book or article, the programs will not detect it unless someone else has used that same quote in one of the papers in the database. They do not compare papers to the volumes of printed

material available for students to use and plagiarize either. Finally, these programs cannot capture the vast number of unsubmitted papers that reside on the hard drives of fellow students and fraternity/sorority members on campuses across the country. Although these limitations are all plainly evident in researching plagiarism products, using a detection program can foster a false sense of security unless they are stressed to faculty during training.

A look at some popular programs

During Fall 2002 and Spring 2003, Wake Forest University investigated three plagiarism detection programs and made a report to the faculty. What follows is a closer look at the services we investigated.

Turnitin

Currently, *Turnitin* seems to be the frontrunner among universities adopting plagiarism detection programs on a large-scale basis. While not inexpensive for large institutions, it clearly offers the most options to faculty. With a departmental or campus license, for example, students can log in and submit their papers themselves.

- **Web site:** <http://www.turnitin.com>
- **Licensing options:**
 - ✓ Individual instructor license (for individual instructors ONLY)
 - ✓ Department license (for individual departments ONLY)
 - ✓ Single campus license (for multiple instructors across departments)
 - ✓ Multi-campus license (for multiple campuses within a system)
- **Cost:** Based on FTE enrollment
- **Services:**
 - ✓ Peer review
 - ✓ Digital archiving
 - ✓ Class management tools

- ✓ Administrator help desk support
- ✓ Internet and database searches
- **Database contents** (from Web site):

“These databases are the result of exhaustive Web searches designed to locate and catalogue information likely to be used for plagiarism. Our ‘Web crawling robots’ retrieve millions of documents from the Internet every day, and by focusing on sites like online paper mills, academic resources, online encyclopedias, and news agencies, we have been able to assemble an in-house version of the Internet tailored to provide the most comprehensive and efficient plagiarism-detection searches available anywhere. In addition to crawling the Internet, we also fingerprint and index papers submitted to us for plagiarism prevention. Our database of submitted papers currently contains over one million documents, and we add thousands of new papers every day. Comparison of new submissions against this database ensures that papers can never be bought, recycled, or traded among students from different years at varying classes and institutions.”
- **Turnaround time:** 3 to 24 hours
- **Report format:**

Called an Originality Report, the *Turnitin* results give each paper a number ranking of “overall similarity” from 1 to 5. It provides links to the items deemed most similar so the instructor can check to see if the similarities are of concern. It then reproduces the text of the paper with items deemed similar to others in the database highlighted and linked to the other sources.

PlagiServe

PlagiServe is a free service that can be used by anyone at any time who wants to set up an account on the Web site. This was one of the programs tested by Wake Forest, and except for a couple of days when the servers were down, our tester found this to be as reliable, if not quite as fast in turn around time, as *Turnitin*. It does,

however, lack the extra features of *Turnitin*, such as being able to set up your own database for a particular class.

PlagiServe received some bad press last year when it was reported that the people who created *PlagiServe* were also responsible for several online paper mills. While this proved to be the case, the creators have abandoned the paper mill business and are now concentrating on their plagiarism detection program services.

- **Web site:** <http://www.PlagiServe.com/>
- **Cost:** Free
- **Services:** Online paper submission and plagiarism detection.
- **Database contents:** (from Web site):

“PlagiServe has created an extensive database of over 150,000 student term papers, essays and cliff notes in order to reduce the usefulness of free as well as paid term paper mills. In other words, if a student has plagiarized his/her manuscript from any out of 100 Paper Mills available in the Internet, PlagiServe will definitely find the source. PlagiServe employs the Internet-wide search when checking the originality of the paper. Searching within more than 1 billion Web pages, PlagiServe uncovers the material used to plagiarize and displays sources to the instructor. PlagiServe’s Web robots have crawled Encyclopedia.com, Refdesk.com, Britannica.com and many other reference sites considered high risk in term of student plagiarism.”
- **Turnaround time:** 12 hours

Report format:
Originality reports are Web-based, with links to matching documents. In this reviewer’s opinion the format is a bit more confusing than *Turnitin*’s, since the user is dealing with split screens and frames.

EduTie

EduTie is the new for-fee version of *PlagiServ* and is based on the same database. It offers more services than the free version.

- **Web site:** <http://www.edutie.com/>

- **Cost:** Based on FTE
- **Licensing options:**
 - ✓ Individual
 - ✓ Departmental
 - ✓ Institutional
- **Services:**
 - ✓ Online paper submission by instructors or students
 - ✓ Document originality detection
 - ✓ Digital archiving
 - ✓ Online grading
- **Database contents:** Same as *PlagiServ*
- **Turnaround time:** 12 hours
- **Report format:**

Much easier to interpret than *PlagiServ*'s format. Ranks papers on a scale from 1 to 5 (original to plagiarized), and then provides the text of the submitted paper with color coding and links for areas that are similar to ones found in the database.

Wake Forest University's pilot evaluation

In order to determine which of these three products best served the needs of our faculty, Wake Forest had two pilots going on simultaneously for the academic year 2002-2003. One was with *Turnitin* alone in the Health and Exercise Science department and allowed the participating faculty to submit their own papers or have students submit papers to the program themselves. Faculty used the programs in First Year Seminars and other writing-intensive classes and there was an overall positive review from the pilot faculty on ease of use and turnaround times for the program.

The other pilot was campus-wide and compared *Turnitin*, *Plagiserv* and *Edutie*. Faculty brought electronic copies of papers to our Teaching and Learning Center Coordinator, who submitted them to all three program services. The faculty later returned to the TLC to view the results. The cumbersome nature of the process kept the number of faculty using the service low, as did a lack of

References

- Argetsinger, Amy. (2001) Technology exposes cheating at U-Va [Electronic version]. *Washington Post*. Wednesday, May 9th 2001, Page A01.
- Foster, Andrea L. (2002) Plagiarism-detection tool creates legal quandaries [Electronic version]. *Chronicle of Higher Education*. May 17th, 2002. Retrieved May 10, 2003, from <http://chronicle.com/free/v48/i36/v48i36/36a03701.htm>.
- Quinlan, Paul. (2002) Bloomfield honor trials finish, 20 found guilty [Electronic version]. *Cavalier Daily Online*. Tuesday, November 26th, 2002. Retrieved May 10, 2003, from <http://www.cavalier-daily.com/CAArticle.asp?ID=14343&pid=943>.

Slobogin, Kathy (2002) Survey: Many students say cheating's OK. *CNN.com*. April 5, 2002. Posted: 1:05 PM EST (1805 GMT). Retrieved May 10, 2003, from <http://www.cnn.com/2002/fyi/teachers.ednews/04/05/high-school.cheating/index.html>.

publicity for the pilot. Many faculty also did not realize that they had to have an electronic copy of a paper for it to be submitted. This caused some anxious moments with OCR software trying to scan in papers that had already been written upon by the faculty member; this process alone kept some faculty from participating. The TLC Coordinator found that for those who did follow the process through, the three programs were very similar in the results they returned—but *Turnitin* was by far the quickest in turnaround time. The added features provided by *Turnitin* and its affordability for a campus the size of Wake Forest, has made it our official choice for the 2003–2004 academic year (and perhaps beyond). Of course, faculty who preferred *Plagiserv* can still use it at any time.

Future issues

The introduction of plagiarism detection programs to the academic environments of our colleges and universities is something that should be studied closely. They bring with them a host of procedural, academic and ethical issues that remain to be worked out. Do faculty, for instance, tell their students that their papers may be submitted to the program? Do faculty require their students to submit their papers themselves—and if so, does this practice promote a culture of suspicion or a leveling of the playing field for all students? Do institutions require faculty to use these programs—and if so, will this create more work for the faculty? Does the fact that some plagiarism detection programs and their faculty users add students' submissions to the database create student privacy or copyright legal complications? This concern in particular has been in the spotlight in the past year and still awaits resolution (Foster, 2002). A list of issues is too long to address properly in a review. But it will grow and change in the coming years, as faculty and their institutions act to counter the impact that the Internet has had on students and their academic activities. ■

A vertical bar on the left side of the page, consisting of a series of horizontal segments in shades of yellow and orange, with a small red diamond at the top.

COPYRIGHT INFORMATION

TITLE: Plagiarism detection programs: A comparative evaluation
SOURCE: Coll Univ Media Rev 9 no2 Spr/Summ 2003
WN: 0310507238009

The magazine publisher is the copyright holder of this article and it is reproduced with permission. Further reproduction of this article in violation of the copyright is prohibited.

Copyright 1982-2003 The H.W. Wilson Company. All rights reserved.