

Parallel Plagiarism Detection Test of TurnItIn, SafeAssign, and Google
Preliminary results, December 5, 2007
Dr. Susan Schorn, Writing Across the Curriculum Coordinator
Department of Rhetoric and Composition

Test Method

I assembled six test essays replicating sources common to rhetoric and humanities classes. I chose sources based on personal experience and the input of instructors I work with in other disciplines. This is not intended to be an entirely representative sample; however, it does very accurately reflect the kinds of sources my students have used and misused (intentionally or not) in my classes over the past ten years. These are the source areas we are most concerned that a plagiarism detector be able to find quickly, accurately, and reliably. These results do not address the problem of collusion between students.

Results

In my judgment, neither SafeAssign nor TurnItIn performed at a level that would justify recommendation to instructors. Google searches consistently out-performed the two proprietary systems in both amount of replicated material found and ease of use. SafeAssign or TurnItIn might outperform Google in collusion detection (not tested here).

Percentage and ratio of finds to misses

Google found more plagiarized sources than SafeAssign or TurnItIn. Google also found the original source more often. There were no results found by SafeAssign or TurnItIn that were not also found by Google.

Performance in test of 23 sources

Software	TurnItIn	SafeAssign	Google
Percentage of sources found	60.76%	43.4%	91.3%
Ratio of sources found:sources missed	14:9	10:13	21:2

Hits and misses are measured in number of sources rather than percentage of text because larger text blocks are easier for search engines to identify. All other things being equal, a system that successfully identifies more, shorter, text blocks is better than one that identifies fewer, but larger, blocks of text.ⁱ

Speed

Google is the fastest of the three systems, providing near-instant results. SafeAssign and TurnItIn require waiting periods of minutes to hours. Of the two proprietary systems, TurnItIn was generally faster.

Depth of results, ease of viewing, and navigability

Google provided superior results in terms of depth and breadth. Each Google search returns multiple “hits,” and allows instructors to assess at a glance how pervasive a text string is on the Internet, where else it has been used, and how recent it is. This information is helpful in determining if plagiarism has indeed taken place. The “streamlining” of results in SafeAssign and TurnItIn simplifies their reports, but the process actually elides important context about the quantity and quality of the source material in question.

It is possible to unpack the “streamlined” results somewhat in TurnItIn, but doing so is time-consuming and still does not result in information as accurate or data-rich as Google results. TurnItIn’s default setting to view matches is “show matches together”—this setting provides only one “hit” per match. The second option (in a drop down menu), “show matches one at a time,” shows additional sources. Instructors would need to have this option explained to them, and then reset the drop-down menu for every viewing of every report. When this second option is used, many results state either “This source is partially hidden by one or more sources in the cumulative report,” or “This source is completely hidden by one or more sources in the cumulative report.” It is not clear what “hidden” means; this would need to be ascertained and explained to instructors.

Google allowed direct viewing of the sources found more often than SafeAssign or TurnItIn. At the time of the test, both proprietary systems privileged student paper results over more original sources. Since the test, TurnItIn states that they have modified their Originality Reports so that “if the same quality match for a section of text is found in both the internet database and the student paper database, we will display the match to the internet in the cumulative view of the Originality Report.”ⁱⁱ Student papers are not usually viewable in either system; therefore, they are of limited value in proving that plagiarism has actually taken place.

Many Web matches that TurnItIn does find are also not viewable by the instructor, because of the limitations of TurnItIn’s crawling and archiving methods. Those that can be viewed are often difficult to read or contextualize.ⁱⁱⁱ

Finally, in almost any attempt to investigate a TurnItIn-identified source more fully, the user eventually reaches a point where he cannot return to the original essay report without closing the window and re-opening the report. This is time-consuming and seems unnecessary; it impedes usability even for the unimpaired and likely poses an even greater obstacle for sight-impaired users.

Usability

Google searching requires basic Internet search skills and little to no special training. The recommended method—selecting word strings containing three to five nouns—is simple.^{iv} The skill of assessing results and fine-tuning searches for harder-to-find sources can be taught in a single 30- to 60-minute session.

In addition to basic user training, SafeAssign and TurnItIn require an additional step for the instructor, who must create an assignment in the SA or TII system. These systems also require students to submit papers through the systems. These are not especially time-consuming steps, but instructors and students must be trained to complete them properly. Mistakes (duplicate assignments, “draft” setting not selected in SafeAssign, papers submitted in the wrong format), which are not infrequent, are time-consuming to correct. As noted above, there are also many individual settings that must be understood and implemented for each of the proprietary systems to be useful to most instructors.

Cost and availability

Google is free and available to anyone with an Internet connection and browser. SafeAssign is free to Blackboard users. TurnItIn requires licensing on a yearly basis.

Document Name	Original Source	TurnItIn	SafeAssign	Google
TeapotDomeTEST.doc	1. George Mason University's History News Network	P ^v	Y	Y
	2. Wikipedia	Y	Y	Y
	3. U.S. Senate	Y	N	Y
	4. GIGA Quotes	Y	N	Y
ShakespearePlag.doc	5. All Free Essays	Y	N	Y
Test essay (Ostrow)	6. Hispanic Online	Y	Y	Y
	7. American Prospect	Y	Y	Y
journalandbookTestEssay	8. Geographical Journal	Y	Y	Y
	9. Applied Animal Behavior Science	N	N	N
	10. A Faith Interrupted (book)	N	N	Y
	11. Tools of Radio Astronomy (textbook)	N	N	Y
	12. A03 Judicial Findings: Liens	P	Y	Y
	13. Babelfish translation of Web page	N	N	N
NightWatchAnalysis	14. Echeat.com	N	P ^{vi}	Y
Race and Culture	15. Counterpunch	N	Y	Y
	16. Newman, David (textbook)	N	P	Y
	17. Encyclopedia Britannica	P	Y	Y
	18. MIT Open Courseware (syllabus)	N	N	Y
	19. Penn State Department of Anthropology	N	N	Y
	20. Rokeach, Milton, 1973. The Nature of Human Values	P	N	Y
	21. Turner, Victor. Symbols in African Ritual	P ^{vii}	N	Y
	22. Turner, Victor. The Forest of Symbols	Y	N	Y
	23. Weber, Max. Economy and Society	P ^{viii}	N	Y

Y = found original source (for Google, indicates find on first page of results)

P = found non-original source; found unviewable source; found only some replicated text from original source; or did not display original source as first result.

N = did not find source

ⁱ With the exception, of course, of text strings so short and/or ubiquitous as to be considered coincidental matches. See note ii, above.

ⁱⁱ Personal email from Michael Bruton, TurnItIn, November 30, 2007.

ⁱⁱⁱ Clicking on highlighted, matched text in an originality report may produce a large unformatted text block (an entire article, for example), in the "match" window. Clicking the small highlighted link "show in web page," which is not obviously placed, opens a new window titled "TurnItIn Direct Source Comparison." The first source I checked this way displayed only the word "unavailable." The second generated a warning, "This web page has been modified since your report was originally generated. Because some or all of the matching text may have been moved or changed, some matches may be missing or incomplete." The page that eventually loaded was an email form that didn't seem to have any bearing on the search in question—it looked like a Web company's customer service page.

^{iv} See Weber-Wulff, Debora (Fachhochschule für Technik und Wirtschaft) <http://plagiat.fhtw-berlin.de/software/>

^v TurnItIn found three sentences of the primary source on George Mason U's site, but attributed two additional sentences from the same site to a student paper.

^{vi} SafeAssign found a single sentence of the essay at echeat.com, but did not flag any of the remaining text. SafeAssign also flagged an (apparently) entirely coincidental match with the sentence fragment "Almost like you could reach out and touch it."

^{vii} TurnItIn found some text embedded in a 13-page German-language PDF at University of Basel, Switzerland.

^{viii} Cannot tell if this match links to the original source. TurnItIn's link is to a subscription-only site.