

Plagiarism detection software: How effective is it?

The first thing to note about plagiarism detection software is that much of it is ephemeral. The available software tends to come and go: new software and websites surface and then disappear. It is difficult to predict how many stable products will emerge.

On the whole, plagiarism detection software can make a useful contribution to minimising plagiarism. The visible use of such software is a strong deterrent to students who are considering plagiarising material. It should be recognised, however, that the software provides no magical answers. Some is expensive; most is time-consuming. No software seems to discriminate between quotations which are properly cited and those which are unacknowledged: what the software detects and notifies is duplication. So reports issued by plagiarism software alert the user to what may appear to be plagiarized material that is in fact appropriately referenced. Manual checking and human judgement are still needed.

Some plagiarism detection schemes require students to submit their work electronically directly to the software company. The company then sends a report on submitted student work to the university. It is important to note that assignment formatting may be lost during the detection process. As a consequence, students must also submit either an electronic copy or hard copy of their work to the university as well – raising the question of how it can be ensured the two versions are identical. If students submit electronic copy, lecturers have to print out the assignment – time-consuming operations in cases where lecturers are assessing large numbers of students.

Some software programs concentrate on comparing the material submitted within a defined group of students; others compare the material submitted with either an in-house data base (compiled cumulatively by the software company), or the web, or both. A number of software plagiarism detection companies offer as part of their service the archiving of student essays, which in turn gives lecturers a specialized in-house data base and, in some cases, the possibility of a secure web environment which can be accessed by students for purposes related to group assessment and peer review.

It is worth remembering, of course, that ‘traditional’ plagiarism — text copied from books rather than downloaded from the Web — may well persist, but cannot be electronically detected.

Plagiarism Detection Software

NAME	FEATURES/ TECHNIQUES	PROS/CONS	COSTS
Copycatch http://www.copycatch.free-serve.co.uk/vocalyse.htm	<p>A UK system which concentrates on comparison within a group of students. The software compares text from work collected by email or on disk using a similarity threshold that will detect essays which are very similar or dissimilar to other class essays by communality of words and phrases.</p>	<p>PROS: The JISC (Joint Information Systems Committee – HEFCE-funded UK organization) gave this software five stars for detection, clarity, value, user-friendliness, speed and reliability. CONS: Detects only collusion among students, and cannot detect material downloaded from the Web</p>	<p>Approx \$700 AUD to purchase software.</p>
Glatt Plagiarism Screening Program (GPSP) http://www.plagiarism.com/INDEX.HTM	<p>Uses the 'fingerprint' method. It exploits the uniqueness of each individual's linguistic patterns – 'cloze' technique. It eliminates every fifth word of a student's paper and replaces the words with a blank which the student is asked to fill in. The number of correct responses is one of the factors considered in the production of a final probability score.</p>	<p>PROS: Useful for detecting plagiarism where the original source material cannot be located. CONS: Students actually have to sit down to a test to fulfil the requirements.</p>	<p>Approx \$580 AUD to purchase software. Additional financial commitment is required for a subscription to the Plagiarism Screening Service to provide the scoring for submitted tests.</p>

NAME	FEATURES/ TECHNIQUES	PROS/CONS	COSTS
<p>Turnitin.com is the user portal for Plagiarism.org.</p> <p>http://www.turnitin.com/</p>	<p>The technology used is called 'document source analysis'. It uses a set of algorithms to make a digital fingerprint of any text document, and then compares it against Internet sources and against an in-house database. Results are compiled into an 'originality report' which colour-codes and underlines text passages showing similarities to other sources, and gives the URLs of the sources.</p>	<p>PROS: Covers a huge range of sources. Offers a digital portfolio service. in which students' work is archived.</p> <p>CONS: The user has to check the report carefully because the software detects correctly-cited material as well as plagiarised material. As in similar programs, formatting is lost in the checking procedure, so essays for marking have to be submitted separately from essays for checking.</p>	<p>A free one-month trial is available. Costs are for subscription rather than purchase and vary according to extent of commitment. The web page provides quotations depending on numbers of classes, numbers of students, and so on.</p>
<p>EVE2 - Essay Verification Engine</p> <p>http://www.canexus.com/eve/index.shtml</p>	<p>Performs searches to find Internet sites with similarities to the submitted text. Produces report underlining text passages possibly plagiarised.</p>	<p>PROS: Tests against wide area of internet.</p> <p>CONS: Each piece of work has to be individually loaded and checked by the lecturer.</p>	<p>Download free for 15 days; purchase for approx \$40 AUD. Each user must purchase a separate copy and licence.</p>
<p>Plagiserve</p> <p>http://www.plagiserve.com/</p>	<p>A system which checks the originality of reports by comparing students' work with its own database and the internet. It provides an originality report that colour codes possibly plagiarised passages and provides direct links to the original source.</p>	<p>PROS: Tests against extensive in-house database and internet searches</p> <p>CONS: Formatting is lost during the checking process, so material has to be handed in separately. Has been claimed to be associated with cheat sites.</p>	<p>Free.</p>

NAME	FEATURES/ TECHNIQUES	PROS/CONS	COSTS
WordCHECK DP http://www.wordchecksyste.ms.com/wordcheck-dp.html	Software which profiles documents by identifying key word use, allowing users to search manually for matching documents based upon word use and frequency patterns. Uses an internal database. Produces report with key-word profiles and word frequency lists.	PROS: Similar system to Copycatch. CONS: Uses only internal database. Manual checking of each piece of work is very time-consuming.	Profiler Basic \$185 (\$115 academic price) Profiler Pro \$570 (\$345 academic price) Users can add "profile capacity" as they go: 2,000 profiles - \$380 5,000 profiles - \$760 10,000 profiles - \$1,540 (All dollars approximate AUD.)
WordCHECK RA http://www.wordchecksyste.ms.com/wordcheck-ra.html	An upmarket package which is aimed at academic research rather than student assignments. Works on the same principles as the DP version.	PROS: As for the DP version. CONS: As for the DP version.	RA Individual Desktop \$380 (\$185 academic price) RA Department Desktop \$1,925 (\$1,347 academic price) plus profile expansion (All dollars approximate AUD.)
Moss http://www.cs.berkeley.edu/~aikenn/moss/html	MOSS is an acronym for Measure of Software Similarity – an internal system at Berkeley developed specifically for computer programming fields.	PROS: Designed with a special focus on computer programming code rather than text. CONS: Limited in scope.	Free but restricted to instructors and staff of computer programming courses. A request must be sent to use.

NAME	FEATURES/ TECHNIQUES	PROS/CONS	COSTS
SIM http://www.cs.vu.nl/~dick/sim.html and http://www.cs.vu.nl/pub/dick/similarity_tester/	Another computer code plagiarism detector. SIM tests lexical similarity in a number of languages including Java, Pascal, Lisp, and Miranda. It detects potentially duplicated code fragments in software projects.	PROS: As for Moss CONS: As for Moss	Available free through the website of the Vrije Universiteit, Amsterdam, where the software was developed.
JPlag http://wwwipd.ira.uka.de:222/	System that finds similarities among multiple sets of source code files. Designed for detecting plagiarism in computer programming but can support plain text as well (although with less satisfactory results).	PROS: The only software that can deal with programming-type work as well as ordinary text. CONS: Limited and less effective in its use with ordinary text.	Free, but an account must be applied for on the website.
Google http://www.google.com	Primarily a search engine and not a plagiarism detector, Google is nevertheless able to detect phrases and can rapidly identify source material from the Internet.	PROS: Quick and free. Google extracts from pdf files, which many search engines cannot do. CONS: Unsystematic, and involves manual entry of strings. Labour-intensive.	Free

Comparison Of Plagiarism Software Features

	Copy catch	Glatt (GPSP)	Turnitin.com	EVE2	Plagi serve	Word CHECK DP	Word CHECK RA	Moss	Sim	JPlag	Google
Checks against Web?	X	X	Y	Y	Y	X	X	X	X	X	Y
Checks against its own data base?	X	X	Y	X	Y	Y	Y	Y	Y	Y	X
Cross-checks with other students' work?	Y	X	Y	X	Y	Y	Y	X	Y	Y	X
Purchase?	Y	Y	X	Y	X	Y	Y	X	X	X	X
Subscription?	X	Y	Y	X	Y	Y	Y	Y	X	Y	X
Free?	X	X	X	X	Y	X	X	Y	Y	Y	Y
Checks ordinary text?	Y	Y	Y	Y	Y	Y	Y	X	X	Y	Y
Checks computer programming?	X	X	X	X	X	X	X	Y	Y	Y	X

	Copy catch	Glatt (GPSP)	Turnitin.com	EVE2	Plagi serve	Word CHECK DP	Word CHECK RA	Moss	Sim	JPlag	Google
Students sit test?	X	Y	X	X	X	X	X	X	X	X	X
Designed for use by students?	X	X	Y	X	X	X	X	X	X	X	Y
Designed for use by teachers?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Instant response?	X	X	X	Y	X	Y	Y	X	Y	X	Y