South African Journal
of Information Management
[DoE accredited]

Secure notes Vol.5(3) September 2003

Grey matter snatchers

S. Berner
sberner@ecognus.com

Plagiarism has been around since time immemorial, but it has now acquired new toys and a new name. Information managers are at the forefront with educators in a concentrated attempt to beat the technology savvy e-cheaters. Will they succeed?

A rose by someone else is not lavender

It seems as if students believe that if content is on-line, not encrypted and their professor is unaware of it, it can be tweaked and pruned a bit and voilà! They have their own product. Why waste time in intellectual pursuit when others' grey matter is lying around for easy picking.

This kind of mentality is becoming a bane of academic and school life. Articles in the *New York Times* (Hafner 2001) and the *Washington Post* (Schulte 2002) portray the situation in the USA. Easy access to information obviously has its dark side, and the shadow is creeping into academia fast. Young people learn the benefits of information tools faster than their teachers, in addition to having more time on their hands to experiment with it. Word of mouth spreads fast, and there have been cases where a whole class downloaded intellectual property from the same vendor on the Internet. Despite the best efforts to teach the ethical use of source material and despite academic institutions' stated policies on academic fraud, it is likely that fraudulent submissions will increase in the future.

Before the advent of the Internet, plagiarism required resourcefulness and skill. Today, even technologically challenged students (and there are very few of those around) have access to huge information resources on-line without having to put too much effort into the search. A few words typed into a Web search engine can lead a student to hundreds of relevant documents. All they must do is 'cut and paste' a few paragraphs from here and a few more from there until an entire paper is 'created'. Students can also find a research paper in one of numerous new journals on-line, copy the entire text, tweak it into a new document and then offer it as an original work. Even ghostwriters have gone global, with Web sites offering professionally written research papers for sale. Some even go as far as to offer a money-back guarantee against detection.

What is at issue here, apart from sheer academic honesty, is copyright. Although rarely mentioned, plagiarizing is akin to illegal burning of music and software CDs, photocopying of whole books or downloading artwork from the Internet to put into one's business pamphlets. All these activities are banned by law and there should be no reason why students (or academics and literati, for they have been guilty too) should not be prosecuted. Academic institutions, however, seem to think otherwise and limit themselves to failing the student or making them re-write a paper. This is despite the fact that plagiarism is actually much easier to detect than illegal software.

The prevalence of the practice is a matter of research. According to Foster (2002), up to 14% of Australian undergraduates may be plagiarizing. Her statistics is based on a research commissioned by six Australian universities from the state of Victoria (Deakin, La Trobe, Monash, RMIT, Victoria Universities and the University of Ballarat) during which 1925 papers were ran through Turnitin, a plagiarism detection software program. In the USA, the numbers are allegedly higher. This could be attributed to a higher ratio of personal computers to users. As many as 72% of high school students and 75% of undergraduates admit to some form of 'serious cheating' (McCabe 2001). On a less drastic note, Scanlon and Neumann (2002) found that only 24,5% of US students actively e-cheat. A similar study among African universities still remains to be undertaken.

Fortunately, not all is lost. Whereas the e-cheaters have their toys, some of which are mentioned below, the academics and university information professionals now have quite a collection of plagiarism-detection software available to them. Whether this will lead to one of the sides prevailing over the other, is questionable, and we may end up watching both sides preening their tools of trade and raising them to even higher standards. Thought policing, after all, has never led to anything.

Toys for e-cheaters

Here follows a short discussion of the most common tools that e-cheaters use.

Paper mills
These sites sell pre-written or custom-made papers or offer certain papers free of charge. Coastal Carolina University Reference Library has a list of quite a few of these mills [http://www.coastal.edu/library/mills2.htm]. Apparently, the latest trend is to set up subject-specific paper mills. In 2001, Stevenson wrote a biting exposé on the quality of the material found on these Web sites. Many of the free research papers have been written and shared by other students. Since paper swappers are often not among the best students, free papers are often of poor quality, both in mechanics and content. Some of the papers are surprisingly old, with citations being no more recent than the 1970s. The papers from commercial paper mills can actually be so good that they smack of professionalism, something rather difficult to find among undergraduate students.

Full-text databases
A lock on the gate has never stopped a burglar. Although it is common practice for full-text databases such as Ebsco, Ingentia or Emerald to clearly state their copyright policies, they do not have means of 'locking' the text so that it cannot be downloaded or copied. Thus, they provide an ever-increasing and updated source of potential material ripe for harvesting by e-cheaters.

On-line language translators
These are absolutely legitimate on-line software applications used for translating text and Web sites. Although their accuracy is never above 60%, they do provide an instant translation of foreign-language texts into a specific target language. E-cheaters use these programs to translate papers in foreign languages into their own language and then tweak the text's style and grammar to submit the paper as original material. In an age where English is becoming the lingua franca, few academics even contemplate a foreign language source as a suspect.

Invisible Web
This category includes the content of sites that are not indexed by search engines and therefore cannot be located by using one. Some magazines, newspapers, reference works, encyclopedias, and subject-specific sites are in this category.

CD-ROMs
There is a wealth of information and reference materials currently available in CD-ROM format, from *Encyclopedia Britannica* to archives of scholarly journals. If this material is not simultaneously available on the Internet, it might prove rather difficult to find the suspect source of plagiarism.

World-Wide Web
A large volume of academic and business research is available on the Web. It only takes typing a good search string into the Advanced Search tool of Google to obtain results. For example, allintitle plagiarism detection software filetype:pdf. These keywords produce full research papers, a wealth of bibliographic information and even contents of full-text databases.

Anti-plagiarism tools review

Before the advent of software and Internet checking methods, teachers and lecturers looked through sources to find the original material. Today, a myriad of software packages and Internet sites can offer help if plagiarism is suspected.

There is a wide range of commercial products and services available for purchase to detect plagiarism. Plagiarism detection software and services can be broadly banded into two groups, namely those designed to detect plagiarism in computer programs and those designed for detecting plagiarism in text-based documents. In this article, the second group is discussed.

Some software programs and services are designed to detect material that is cut and pasted from the Internet, while others detect instances of identical or very similar submissions. Some services have the facility to compile databases and so build up a selection of assignments and material that has been purchased from paper mills and essay banks. In this way, a second assignment that is submitted is matched against the captured material. Some of the services combine several features and offer solutions for detecting different forms of plagiarism. A few of the wider used services are reviewed below. The name of the company is followed, in brackets, by the name of the product and the company's Web address.

- Digital Integrity (Findsame, http://www.digital-integrity.com): Designed to detect material cut and pasted from the Internet. This service is based on MOSS technology. It seems more geared towards the needs of the academic community.
- CaNexus (EVE2, http://www.CaNexus.com): Designed to detect material cut and pasted from the Internet. EVE2 is a search engine that performs complex searches to find material from the Internet. It does not compile material into a database and so cannot directly compare texts. It does not trace documents that are not in HTML format. Therefore, it does not trace material copied from discussion boards. EVE2 is available to download for a free 15-day trial of the full software program.
- iParadigms (Turnitin, http://www.turnitin.com): Turnitin is a Web-based subscription service that offers simultaneous Internet plagiarism and collusion detection. Turnitin detects material copied from the Internet, as well as collusion between students through cross-checking of submitted essays against one another and against an in-house database of texts. When an essay is submitted for checking, it is added to the database for future reference. Turnitin can trace essays to paper mills, but cannot find material copied from discussion boards unless uploaded separately. A free restricted trial account is available that allows submission of five manuscripts over a period of 30 days. The trial account does not give access to the Turnitin database.
- CLF Software (CopyCatch, http://www.CopyCatch.freeserve.co.uk): CopyCatch detects collusion between students by comparing submitted documents and calculating

the proportion of words held in common. Comparison at phrase level is also possible. CopyCatch does not detect plagiarism from the Internet.

- WordCHECK Systems (WordCheck, http://www.WordCHECKsystems.com): WordCHECK detects collusion between students by matching key word profiles between a submitted document and documents kept in an internal archive. It does not detect plagiarism from the Internet. WordCHECK does not allow profile matching at phrase level. A profile matching add on to WordCHECK can be bought at an additional cost (Hughes, Brown, Jakobson, Philpot, Dwight-Moore, Jarrett, Grainger and Short 2002).
- Integriguard (Howoriginal and Paperbin, http://www.integriguard.com): Compares submissions against a database of other papers and Web sites. It offers two ways of combating plagiarism. Its HowOriginal.com site works on a submitted paper, comparing it to its database of papers as well as to Web searches and provides a report showing any matching phrases it finds.
- Glatt Plagiarism Services (Glatt Plagiarism Screening, www.plagiarism.com): An interactive program that uses the student's own writing style. This software is especially useful for detecting plagiarism when the source is another student, a purchased paper or a book. It offers computer software to teach students to avoid plagiarism and also provides students with the opportunity to self-test for plagiarism.
- Plagiserve (PlagiServe, www.plagiserve.com): This on-line service checks one paper in a class against other papers in the same class (clearly, all papers must be uploaded), as well as against papers available on paper mills. Plagiserve claims to have the largest database to check papers against. One can also check fragments of papers with this free service. However, read the concerns raised about this site at http://plagiarism.phys.virginia.edu/links.html before using it.

For an excellent review of anti-plagiarism software, refer to Bull, Collins, Coughlin and Sharp (2001) and Culwin and Lancaster (2000).

Role of information professionals in combating plagiarism

As libraries increase their number of full-text resources, such as electronic journals, Web sites and periodical databases (e.g. InfoTrac), the need to educate users about the ethical use of information becomes more important. It is obvious that students are in great need of guidance on how to use information ethically and legally. Instructional sessions with information professionals should include direct information about plagiarism and its consequences along with practical steps students can take to avoid the risk of plagiarism in their research assignments.

Information professionals are by nature of their training the best-positioned academic staff to deal with plagiarism. Their knowledge of information resources, mastery of search techniques and access to a wide variety of databases puts them ahead of teachers who may or may not be as competent in on-line information acquisition. In addition to actively cooperating with educators in checking for plagiarized material, information professionals should also be involved in educating both teachers and students. Information professionals should also work with teachers and lecturers to re-examine their curriculum to identify points and places where plagiarism should be discussed with students. Although in the past the librarian's role on campus was limited to point-of-use aides, reference interviews and instructional classes, information professionals must now actively seek new roles that will open regular dialogue with students about the ethical use information. Information professionals can take an active part in educating students in areas of research and can provide teachers with tools and alerts to what is lurking in the deep murky cyber-waters of e-cheating. Auer and Krupar (2001) mention a few methods to achieve this:

- Inform faculty about paper mills, software and the Internet: Information professionals should begin a dialogue with faculty about plagiarism and provide information about Web sites and software. Once plagiarism is suspected, the librarian can help through both traditional and technology-oriented methods.
- Instruction: Instructional sessions are the perfect method for providing students with information about the ethical use of Web sources and full-text articles in their research. Information professionals have an ethical obligation to teach bibliographic citation methods and strategies on how to best avoid plagiarism of especially Internet sources.
- Handouts: Print and on-line guides must be readily available for students to find information on their own. Therefore, information professionals must provide students and faculty with information, in various formats, about citing on-line information. Both print handouts and Web pages can give students information about how to use various citation styles and where to find more information on this issue. Handouts are particularly useful since they can be used in any setting and students can write notes directly on them for future reference.

Caveat emptor

Depending on where you are and which privacy and copyright laws apply to your situation, you should be careful that the plagiarism-combating technology (PCT) that you are using does not clash with your students' legal rights. Turnitin has been queried by US legal eagles on its database of submitted papers, which potentially infringes on the student's automatic copyright according to US law (Foster 2002).

Apart from legal infringement, the other issue with plagiarism detection software is the breach of trust it creates between the students and their teachers. Quite a few universities in the US are refusing to use the software and at academic institutions where it is in use the number of subscribed academics is small (Beck 2002).

Another thing to be weary of is that most software, when used for the first time, picks up everything in the text, including permissible quoting, footnotes and paraphrasing. Even copies of one's own work that is available on-line can be picked up (Dehnart 1999). Do not judge a paper by the percentage the software gives it. Instead, read the paper and judge for yourself whether the text is 'legal' or not. Many of the software packages actually highlight the 'suspect' text in different colours – Turnitin has a different colour for different 'suspected' sources.

In the end, the onus is on academics to be reasonable. Students are there to learn. An essential part of learning is becoming familiar with the ideas and writings of other people. Examples of other writing and research are widely provided and recommended. Reading lists specify specific books, articles and Web sites. We should therefore not be surprised if students produce mosaics or patchworks of the writings of others. The best way to combat plagiarism is good pedagogy: being fair but firm with the students, making them aware that they can easily be found out, encouraging creativity and radical thinking, teaching students how to research and think analytically, and setting clear boundaries as to what is permissible and what the consequences are. Although I disagree with Howard (2002) that we should ignore the whole issue, as a previous educator I would like to stress that breeding integrity and pride in one's own abilities and work is, as Howard rightly implies, the best anti-plagiarism software.

References

Auer, N. and Krupar, E. 2001. Mouse click plagiarism: the role of technology in plagiarism and the librarian's role in combating it. *Library Trends* 49(3):424-429.

Beck, H. 2002. McGill refuses to use plagiarism detectors. *The McGill Daily*. [Online]. Available WWW: http://www.previewmysite.com/mcgilldaily.com/archives/january102002/news05.html (Accessed 31/07/03).

Bull, J., Collins, C., Coughlin, E. and Sharp, D. 2001. *Technical review of plagiarism detection software report*. Luton, Bedfordshire: University of Luton.

Culwin, F. and Lancaster, T. 2000. *A review of electronic services for plagiarism detection in student submissions*. Paper presented at the Learning and teaching support network for information and computer sciences first annual conference. 23–25 August. Edinburgh. [Online]. Available WWW: http://www.ics.ltsn.ac.uk/pub/conf2000/Papers/Culwin.pdf (Accessed 1 August 2003).

Dehnart, A. 1999. *The Web's plagiarism police*. [Online]. Available WWW: http://www.salon.com/tech/feature/1999/06/14/plagiarism/print.html (Accessed 1 August 2003).

Foster, A. 2002. Plagiarism-detection tool creates legal quandry. *The Chronicle of Higher Education* 48(36):A37. [Online]. Available WWW: http://chronicle.com/free/v48/i36/36a03701.htm (Accessed 31 July 2003).

Foster, A. 2002. Up to 14% of Australian university students may be plagiarizing from Web, study suggests. *The Chronicle of Higher Education*. [Online]. Available WWW: http://chronicle.com/free/2002/11/2002112001t.htm (Accessed 1 August 2003).

Hafner, K. 2001. Lessons in Internet plagiarism. *New York Times*. [Online]. Available WWW: http://www.nytimes.com/learning/teachers/featured_articles/20010629friday.html (Accessed 31 July 2003).

Hughes, G., Brown, S., Jakobson, M., Philpot, C., Dwight-Moore, P., Jarrett, N., Grainger, T. and Short. B. 2002. R*eport on the viability of CopyCatch plagiarism detection software*. London: University of East London.

McCabe, D. 2001. *Students cheating in American high schools*. [Online]. Available WWW: http://www.academicintegrity.org/cai_research.asp (Accessed 1 August 2003).

Scanlon, P. and Neumann, D. 2002. Internet plagiarism among college students. Journal of *College Student Development* 43(3): 374–385.

Shulte, B. 2002. Cheatin', writin' and 'rithmetic. *Washington Post*, pp. W16. [Online]. Available WWW: http://hawks.vwh.net/english/robb/documents/Cheatin,%20Writin,%20Rithmetric.htm (Accessed 31 August 2003).

Stevenson, S. 2001. Adventures in cheating: a guide to buying term papers. *Slate*. [Online]. Available WWW: http://slate.msn.com/id/2059540/ (Accessed 31 August 2003).

Further reading

Buranen, L., Roy, A. and Lunsford, A. (eds.) 1999. *Perspectives on plagiarism and intellectual property in a postmodern world.* Albany (NY): State University of New York.

This book offers a wealth of thinking about the complex and often contradictory definitions surrounding the concepts of plagiarism and intellectual property. The authors show that plagiarism is not nearly as simple and clear-cut a phenomenon as we may think. It contains a variety of analytical approaches to examining the phenomenon: a content analysis of writing

instruction textbooks, historical studies of academic misbehavior in American colleges, surveys of policy statements oriented toward students, case studies, and what might be seen as a law review article.

Decoo, W. 2002. *Crisis on campus: confronting academic misconduct*. Cambridge (MA): MIT.

The book pays special attention to plagiarism as one of the most frequent but also most complex forms of academic misconduct. It analyses the various degrees of possible plagiarism, detection techniques, challenges in proving plagiarism and denial tactics. It gives valuable advice on how to report and handle cases of alleged plagiarism, both by students and by professionals.

La Thrope, A. and Foss, K. 2000. *Student cheating and plagiarism in the Internet era: a wake-up call.* Westport (CT): Libraries Unlimited.

Put a stop to high-tech and more traditional low-tech forms of cheating and plagiarism. Also, learn to recognize the danger signs for cheating and how to identify material that has been copied. Sample policies for developing academic integrity, reproducible lessons for students and faculty, and lists of helpful online and print resources are included. A must-read for concerned educators, administrators and parents.

Mallon, T. 2001. *Stolen words*. Fort Washington (PA): Harvest.

Mallon sees the Internet as opening up vast new opportunities for the enterprising plagiarist and making the job of those who would thwart them a great deal more difficult. Mallon shows how major literary figures such as Laurence Sterne, Coleridge and de Quincey infused their works with ample unattributed borrowings.

Related Web sites
Plagiarism, Cyberplagiarism and Possible Solutions from Southwest Missouri State University Libraries (http://library.smsu.edu/LIS/workshops/wplagiarism.shtml) is a good place to start. The site provides a list of available plagiarism detection tools, commercial paper vendors and a few good articles.

Plagiarism: keeping up with the cheats – a bibliography by John Royce. (http://vm.robcol.k12.tr/~jroyce/plagbibl2.html) is rich in links despite the rather unappealing interface.

The 2000 CELT Conference at California State University (http://library.shastacollege.edu/papermills.htm) has material on paper mills and plagiarism – if you can stand their choice of fonts. Check out their 'further reading' link.

About the author
Sam Berner (B.Ed., Dipl. LIS, Postgraduate Diploma in Information Management) is a principal of the company ECognus (Brisbane, Australia). She is a knowledge management consultant, assisting small to medium enterprizes to benefit the most from their intellectual assets. ECognus also provides services in the area of tailored software applications and the digitization of business processes.

Disclaimer

Articles published in SAJIM are the opinions of the authors and do not necessarily reflect the opinion of the Editor, Board, Publisher, Webmaster or the Rand Afrikaans University. The user hereby waives any claim he/she/they may have or acquire against the publisher, its suppliers, licensees and sub licensees and indemnifies all said persons from any claims, lawsuits, proceedings, costs, special, incidental, consequential or indirect damages, including damages for loss of profits, loss of business or downtime arising out of or relating to the user's use of the Website.

ISSN 1560-683X

Published by InterWord Communications for the Centre for Research in Web-based Applications, Rand Afrikaans University