

---

# Plagiarism issues for higher education

by Fintan Culwin and Thomas Lancaster, School of Computing, Information Systems and Mathematics, South Bank University

*Academic institutions are finding they have to operate under a pro-active anti-plagiarism policy, where plagiarism is actively sought out as a serious breach of acceptable academic behaviour. This paper considers the reasons that institutions need such a policy and the issues they should be aware of when implementing one.*

## Introduction

The ever-growing amount of information on the Internet and the ease with which students have access to this information has produced an increasing temptation for students to download Web material, possibly modify it and hand it in as their own unaided work. Taking the words or ideas of another person and using them without proper acknowledgement is a commonly accepted definition of plagiarism. *Student plagiarism* is further defined as plagiarism with the intent of gaining academic credit. Academic institutions should be concerned about this since student plagiarism appears to be on the increase<sup>1, 10, 14</sup>.

This paper considers why academic institutions are having to employ already scarce resources to combat plagiarism. It also looks at the issues that institutions should be aware of when trying to operate a pro-active anti-plagiarism policy.

## The problem of student plagiarism

Students intending to gain academic qualifications are expected to demonstrate appropriate levels of attainment and ability through coursework and examinations. This requires students to produce

submissions that meet a given *assignment specification* which is then marked by a tutor to confirm that the work reaches the required standard. In many, if not the majority, of institutions students are also required to confirm that the submission is the result of their own, unaided work. Students who falsely give this declaration are playing a part in reducing the value of the qualifications awarded by the academic institution. Knowing that other students are cheating, but are not being punished for it, can be infuriating to other students, who may themselves be discouraged from putting appropriate effort into their own submissions.

The Franklyn-Stokes and Newstead studies in 1995, taken as the British higher education baseline, showed that over half of all students are willing to confess to having plagiarised at some stage of their academic course<sup>9</sup>. This included not only *intra-corporal plagiarism*, copying from other students on the same course, but also *extra-corporal plagiarism*, such as close paraphrasing from a book or submitting an assignment produced in response to an identical specification in a previous year. Such definitions were standardised by Culwin and Lancaster in their plagiarism taxonomy<sup>5</sup>. Newstead, Franklyn-Stokes and Armstead found that students cheated mainly due to time pressure, or because they didn't believe that they could get marks as high as they wanted without resorting to cheating<sup>13</sup>. They also found that a significant proportion of students cheated because they knew of other students cheating and hence believed it was the norm.

The current proportion of students who are plagiarising is expected to be higher than in the Franklin-Stokes and Newstead study. Their figures were produced before the growth of the Web, which provides students with easy access to vast amounts of information that can be copied. This growth of *Web plagiarism* has been discussed at length in the literature<sup>1, 10, 14</sup>. Some students assemble a submission by amalgamating extracts taken from a number of different sites. Others copy assignments directly from sites supplying free pre-written essays<sup>17, 22</sup>. Any Americanisms in these essays can be swiftly corrected with the aid of a UK spell checker and cultural references altered with a minimum of effort.

Critics have attempted to discuss why Web plagiarism is a particular problem and how it can be countered. Ryan gives examples of pointers from her own experience and looks at how a tutor can

use Web search engines to attempt to find out where the plagiarism has been sourced from<sup>14</sup>. Gajadhar discusses the belief of some students that material on the Internet is 'free' and can be reused without citations and concludes that many institutions are ignoring the problem because they don't know how to cope with it<sup>10</sup>. Austin and Brown discuss how assignment specifications can be set that minimise the ways in which students can cheat and how students can be made aware that plagiarism is unacceptable<sup>1</sup>. A search for plagiarism on the Internet will also throw up a large number of Web sites that offer advice and give their solutions to the problems<sup>11, 19, 20</sup>. All comment about the magnitude of the problem and how it is an issue worth serious attention.

## Collapsing a cheating culture

Much of the rigorously researched evidence for the extent of plagiarism is pre-Web and hence obsolete. But the problem is receiving an increasing amount of attention from different quarters. The press seem to find a story on plagiarism every week<sup>3, 4, 16</sup>. Academics discussing the problem say they routinely come across plagiarism cases. Academic organisations are finding it necessary to be seen to be taking plagiarism seriously. The Joint Information Systems Committee (JISC) has held workshops<sup>2</sup> and started to fund anti-plagiarism research. The Committee of Vice Chancellors and Principals (CPVP) is also investigating the extent and impact of plagiarism as well as technology to detect it. The sudden rise in the number of journal and conference submissions also suggest that plagiarism is a hot issue and most of these sources suggest that it is both prevalent and endemic<sup>12, 15</sup>.

Institutions seem to be struggling to update their policies to be seen to be taking plagiarism seriously in light of this interest. Policies vary from *reactive* policies, where plagiarism is treated as a serious academic offence when it is found, but not actively sought out, to *pro-active* policies, where an effort is made by institutions to find and reduce plagiarism by technical and other means. Many institutions are finding it necessary to replace a reactive policy with a proactive one to break the seemingly growing culture that thinks cheating in our academic institutions is acceptable.

A pro-active policy involves standard plagiarism prevention measures. These range from not issuing the same assignment specification year after year,

researching the subject area to check that model essays are not prevalent on the Internet and requiring drafts of submissions to minimise the opportunities for students to cheat<sup>11</sup>.

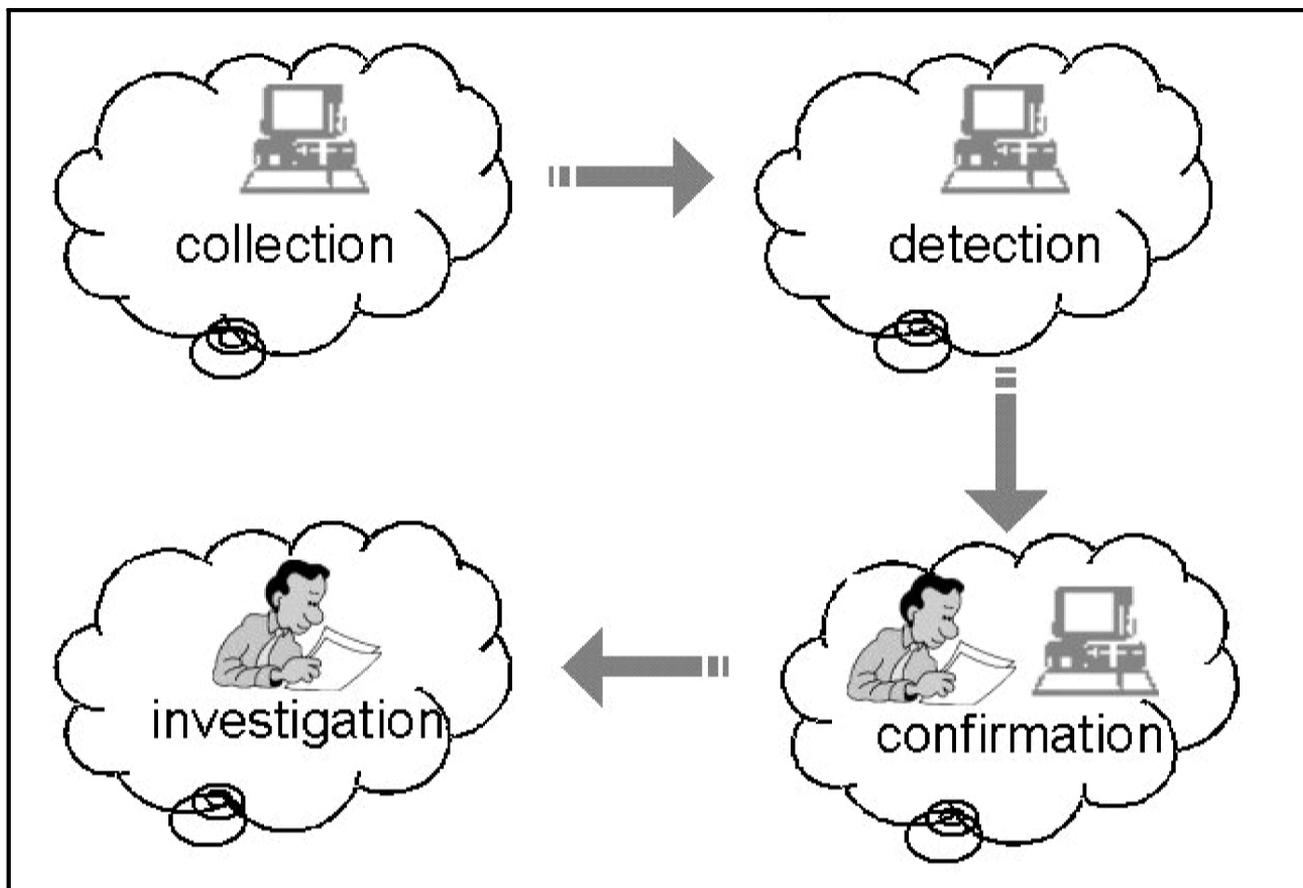
Institutions are having to employ increasingly technical solutions in order to implement their proactive anti-plagiarism policies, including using Web-based plagiarism detection services. Culwin and Lancaster reviewed a number of these, finding each to provide much the same service, with sections of a student submission hyperlinked to Web pages containing similar text<sup>7</sup>. A smaller study by Denhart reached a similar conclusion<sup>8</sup>. The main problem with such solutions was seen to be that they were economically unfeasible. The market leader plagiarism.org charges \$1 for each paper that is submitted to it. The site was also found to be suitable for accepting submissions in plain text only and not the word processing formats most students use. It did not make the degree of similarity between two submissions clear to the user. However plagiarism.org<sup>21</sup> is being trialled by a number British institutions in a current JISC project.

Institutions with only a reactive anti-plagiarism policy have access to a similar service. Findsame.com<sup>18</sup> offers limited free detection for single submissions, via copy and paste into a text box, which can be useful to find Web plagiarism. This very much places the burden of work on the tutor responsible for marking the student submission. For a complete check they would have to put all the submissions of their students through the detection engine on an individual basis and check all the results. However, this service is valuable to tutors and much more useful than individually submitting lines of a submission into a search engine in an attempt find matching text.

## Four-stage plagiarism detection process

The Four-Stage Plagiarism Model, as shown in Figure One, is designed to illustrate how an automated pro-active approach to plagiarism detection can be implemented.

In the *collection stage*, students submit their work to the system, usually via a Web front end. Next, in the *detection stage*, the collected work is run through a detection engine, which produces a list



**Figure 1 – Four-stage plagiarism detection process**

of those student submissions, or pairs of submissions, which appear to be the most similar. The human led stage of *confirmation* follows, where a human manually verifies that the similarity reported represents plagiarism. The system may be erroneously reporting similarity, this is known as a *false hit*. The similarity may also have come about by two people legitimately quoting from the same cited sources, which would not constitute plagiarism. Any similarity still thought to be plagiarism is passed on to the further *investigation* stage which might result in a penalty.

The main issues in the collection stage all stem from the fact that it is *logistically complex*. It is necessary to know who is going to submit work and what the work is that they are submitting. Any submission system has to take into account that the work will have deadlines which can differ from student to student should the circumstances arise. Work may need to be collected in a variety of formats, e.g. Rich Text or MS Word formats for a report type question, or a standard text file for programming source code. Whatever format the submissions are collected in, they need to be converted into plain text for machine analysis. Tutors will need to be informed as to which

students work has and has not been collected so they can follow this up if necessary.

Where a collection system completely replaces a paper based submission there is also the necessity to distribute the collected work amongst the markers, either in printed or electronic form. The machine collection also has the additional advantage of decreasing the workload of the department's clerical staff. The submissions need to be stored securely, complying with Data Protection Act requirements, which also mean that students must be clearly be informed why their work is being collected on-line and how the information from processing them will be utilised.

Detection is the *computationally intensive* stage, since a large number of student submissions have to be compared against one another to detect intra-corporal plagiarism, and against external sources to find extra-corporal plagiarism. The main problem is to find what possibly disguised parts of submissions are stored where. In addition to locating true hits, detection systems also have to be non-capricious by attempting to ensure that they avoid *missed pairs*. A missed pair is a pair of submissions with significant similarity, but which

the detection engine does not flag as such. These considerations have to be balanced against the storage and computational requirements of the detection engine. Additionally, submissions could be processed individually as they come in or batch processed, either when a whole set or *corpus* of submissions is ready or at regular time intervals.

Tutors must confirm any detected similarity in what could be said to be a *judgementally complex* stage of the detection process. To be competent tutors have to ensure that they are consistent, non-capricious and coherent. This suggests that anyone who is involved in the plagiarism detection process would require training and would require a good knowledge about what undue similarity actually is.

The final stage, investigation, is *jurisprudentially complex*. If a submission has been found to contain evidence suggesting plagiarism a decision has to be made whether to exonerate or penalise the student who submitted it. The students have a right to present circumstances that mitigate the cheating and possibly to appeal against any penalty. This process, in the UK, must ensure that students are presented with the evidence that suggests an academic offence has been committed and given

the opportunity to defend themselves. Moreover the people involved in collecting the evidence must be distinct from those making the judgement. If this due process is not followed it is possible that a leave to appeal to the civil courts could be obtained and subsequently damages awarded against the institution.

Any department considering the introduction of a pro-active policy must take care to compare the costs incurred, primarily valuable staff time, with the benefits obtained, mainly ensuring that only students doing their own work receive academic credit.

The issues presented in this article have been informed by the authors own Visualisation and Analysis of Similarity Tool (VAST), designed to vastly decrease the amount of work tutors need to do to confirm and investigate possible plagiarism in the human led stages of the Four-Stage Plagiarism Detection Process<sup>6</sup>. Figure Two shows VAST being used to identify plagiarism in student submissions. The pronounced areas in the graphic on the left represent excessive undue similarity; the similarities in the highlighted area are shown in the windows on the right.

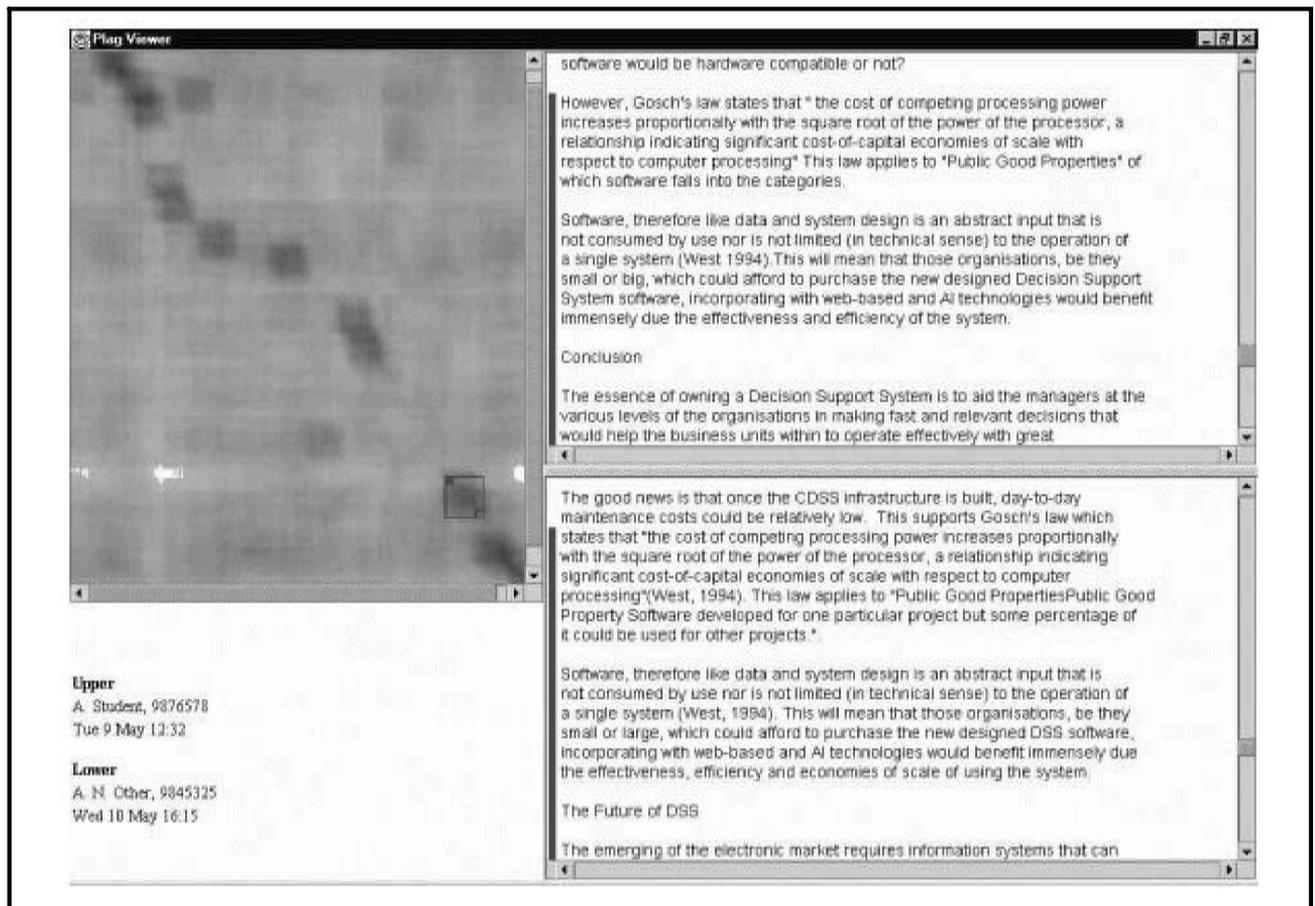


Figure 2 – VAST used to investigate student submissions

The current JISC project, primarily investigating plagiarism.org, but also producing a *Which* style report on other detection engines is investigating some collection issues. However, it is not concerned about the confirmation and investigation stages. This is despite these being, possibly, the most crucial stages in the whole procedure and certainly the most time consuming for a human investigator.

The Quality Assurance Agency (QAA) Code of Practice states that institutions should ensure that assessment is conducted with rigour, fairness and due regard for security, whilst preventing fraudulent activities, such as impersonation and students submitting work that is not their own. Institutions wanting to take the QAA code seriously need to have a pro-active anti-plagiarism policy.

## Conclusions

Many stakeholders are convinced that plagiarism is a serious academic problem and institutions will have to demonstrate that they are tackling it. The use of automated systems is expected to become more and more common, but they need to be informed by a greater understanding of how and why students cheat.

There is certainly a lack of consistency across the whole Higher Education sector. Some institutions have a pro-active anti-plagiarism policy, some a reactive policy and a few still claim, but cannot prove, that none of their students cheats. The standards across the sector need to be equal, so that students cannot assume that they can cheat by moving to a more lenient institution. The QAA guidelines alone are not sufficient for this and further guidance on plagiarism policies, plus guidance on issues of legality, such as jurisprudence, the Data Protection Act and copyright, is urgently needed.

The best possible solution would simply see the culture of cheating eliminated; few people think this is possible, even if a combination of technical solutions placed alongside carefully designed courses and assignment specifications can reduce it to a minimum. Institutions that do act proactively run the risk of reducing their student numbers and income in the short term. Hopefully, publicity of the value of their awards long-term should make up for it.

## References

1. Austin M. & Brown L., *Internet Plagiarism: Developing Strategies to Curb Student Academic Dishonesty*. *The Internet and Higher Education*, 2, 1, p21-33 (1999).
2. Barrie J. *Presentation on iParadigms and plagiarism.org*. At JISC Plagiarism Workshop, South Bank University, London (July 7, 2000).
3. Clare J., *Computer plagiarism 'threatens the value of degrees'*. Daily Telegraph 3/7/2000, Available online at <http://www.telegraph.co.uk/et?ac=004228431730477&rtmo=VDw3wDqK&atmo=rrrrrrrq&pg=/et/00/7/3/ncopy03.html> (2000), [Accessed 26/1/2001].
4. Cramb A., *University withholds 90 exam results over 'Internet cheating'*. Daily Telegraph 10/7/1999, Available online at <http://www.telegraph.co.uk/et?ac=004228431730477&rtmo=VDw3wDMK&atmo=rrrrrrrq&pg=/et/99/7/10/ncheat10.html> (1999), [Accessed 26/1/2001].
5. Culwin F. & Lancaster T., *A Descriptive Taxonomy of Student Plagiarism*. Awaiting publication, available from South Bank University, London (2000).
6. Culwin F. & Lancaster T., *Visualising Intra-Corpall Plagiarism*. Accepted for IEEE Information Visualisation 2001, 25-27 July, London, UK, available from South Bank University, London (2000).
7. Culwin F. & Lancaster T., *A Review of Electronic Services for Plagiarism Detecting in Student Submissions*. 8<sup>th</sup> Annual Conference on the Teaching of Computing, organised by the LTSN Centre for Information and Computer Sciences (2000).
8. Denhart A., *The Web's Plagiarism Police*. Available online at <http://www.salon.com/tech/feature/1999/06/14/plagiarism/index.html> (1999) [Accessed 26/1/2001].

9. Franklin-Stokes A. & Newstead S., *Undergraduate Cheating: Who Does What & Why?* *Studies in Higher Education*, 20, 2, p159-172 (1995). <http://www.guardianunlimited.co.uk/Archive/Article/0,4273,4032413,00.html> (2000), [Accessed 26/1/2001].
10. Gajadhar J., *Issues in Plagiarism for the New Millennium: An Assessment Odyssey*. Available online at <http://ultibase.rmit.edu.au/Articles/dec98/gajad1.htm> (1998) [Accessed 26/1/2001].
11. Harris R., *Anti-Plagiarism Strategies for Research Papers*. Available online at <http://www.vanguard.edu/rharris/antiplag.htm> (2000) [Accessed 26/1/2001].
12. Lancaster T. & Culwin F., *Towards an Error Free Plagiarism Detection Process*. Accepted for ITiCSE 2001, 25-27 June 2001, Canterbury, UK, available from South Bank University, London (2001).
13. Newstead S., Franklyn-Stokes A. & Armstead P., *Individual Differences in Student Cheating*. *Journal of Educational Psychology*, 88, 2, p229-241 (1996).
14. Ryan J. J. C. H., *Student plagiarism in an online world*. Available online at [http://www.asee.org/prism/december/html/student\\_plagiarism\\_in\\_an\\_onlin.htm](http://www.asee.org/prism/december/html/student_plagiarism_in_an_onlin.htm) (1998) [Accessed 26/1/2001].
15. Sheard J., Dick M. & Markham S., *Is it Okay to Cheat? The Views of Postgraduate Students*. Accepted for ITiCSE 2001, 25-27 June 2001, Canterbury, UK (2001).
16. Sutherland J., US students log on to the Internet for a cheat's charter. *Guardian*, 22/6/2000, Available online at <http://www.guardianunlimited.co.uk/Archive/Article/0,4273,4032413,00.html> (2000), [Accessed 26/1/2001].
17. *Cyber Essays*. Available online at <http://www.cyberessays.com> [Accessed 26/1/2001].
18. *Digital Integrity Search*. Available online at <http://www.findsame.com> [Accessed 26/1/2001].
19. *Plagiarism and Anti-Plagiarism*. Available online at <http://newark.rutgers.edu/~ehrllich/plagiarism598.html> [Accessed 26/1/2001].
20. *Plagiarism: What It is and How to Recognize and Avoid It*. Available online at <http://www.indiana.edu/~wts/wts/plagiarism.html> [Accessed 26/1/2001].
21. *Plagiarism.org*. Available online at <http://www.plagiarism.org> [Accessed 26/1/2001].
22. *School Sucks*. Available online at <http://www.schoolsucks.com> [Accessed 26/1/2001].

#### Contact details

Fintan Culwin  
Email: [fintan@sbu.ac.uk](mailto:fintan@sbu.ac.uk)

Thomas Lancaster  
Email: [lancaste@sbu.ac.uk](mailto:lancaste@sbu.ac.uk)

School of Computing, Information Systems and Mathematics  
South Bank University  
London, SE1 0AA  
<http://cise.sbu.ac.uk>